Toxoplasma gondii Toc75 functions in import of stromal but not peripheral apicoplast proteins

Lilach Sheiner[1,2], Justin D. Fellows[1], Jana Ovciarikova[2], Carrie F. Brooks[1], Swati Agrawal[1], Zachary C. Holmes[1], Irine Bietz[3], Nadine Flinner[4], Sabrina Heiny[3], Oliver Mirus[4], Jude M. Przyborski[3] and Boris Striepen[1]

[1] Center for Tropical and Emerging Global Diseases & Department of Cellular Biology, University of Georgia, 500 D.W. Brooks Drive, Athens, GA 30602, USA

[2] Wellcome Trust Centre For Molecular Parasitology, Institute of Infection, Immunity & Inflammation, College of Medical, Veterinary & Life Sciences, Sir Graeme Davies Building, University of Glasgow, 120 University Place, Glasgow G12 8TA

[3] Department of Parasitology, Faculty of Biology, Philipps University Marburg, Marburg, Germany

[4] Molecular Cell Biology of Plants, Biocenter N200, 3. OG, Max-von-Laue-Str. 9, 60438 Frankfurt

Running title: *Toxoplasma* Toc75 in apicoplast protein import

Keywords: Omp85, Toc75, Sam50, *Toxoplasma*, *Plasmodium*, Apicomplexa, protein trafficking, apicoplast, complex plastid.

Corresponding author:
Lilach Sheiner. Wellcome Trust Centre For Molecular Parasitology, University of Glasgow, 120 University Place, Glasgow G12 8TA. Phone: +44(0) 141 330 4797 Lilach.Sheiner@glasgow.ac.uk

Synopsis: Protein targeting to plastids and mitochondria of parasites relies on an elaborate system of signals and machinery. We describe *Toxoplasma* and *Plasmodium* Toc75 and Sam50 proteins. TgToc75 is found to mediate stromal but not peripheral apicoplast protein import and to be essential for parasite growth and plastid maintenance

Abbreviations: ATc (Anhydrous tetracycline), POTRA (polypeptide-transport-associated), OMP85 (outer membrane proteins of 85kDa), TIC/TOC (translocons of the inner/outer chloroplast membrane), SIMM (second innermost membrane), PPC (periplastid compartment), Sam50 (sorting and assembly machinery of 50kDa).

1

2 ## *Toxoplasma gondii* Toc75 functions in import of stromal but not
3 peripheral apicoplast proteins

4
5 Lilach Sheiner[1,2], Justin D. Fellows[1], Carrie F. Brooks[1], Swati Agrawal[1], Zachary C. Holmes[1],
6 Irine Bietz[3], Nadine Flinner[4], Sabrina Heiny[3], Oliver Mirus[4], Jude M. Przyborski[3] and Boris
7 Striepen[1]

8

9 [1] Center for Tropical and Emerging Global Diseases & Department of Cellular Biology, University of
10 Georgia, 500 D.W. Brooks Drive, Athens, GA 30602, USA

11 [2] Wellcome Trust Centre For Molecular Parasitology, Institute of Infection, Immunity & Inflammation,
12 College of Medical, Veterinary & Life Sciences, Sir Graeme Davies Building, University of Glasgow, 120
13 University Place, Glasgow G12 8TA
14 [3] Department of Parasitology, Faculty of Biology, Philipps University Marburg, Marburg, Germany
15 [4] Molecular Cell Biology of Plants, Biocenter N200, 3. OG, Max-von-Laue-Str. 9, 60438 Frankfurt
16
17 **Correspondence**: Lilach.Sheiner@glasgow.ac.uk
18
19 **Running title:** *Toxoplasma* Toc75 in apicoplast protein import
20
21 **Keywords**: Omp85, Toc75, Sam50, *Toxoplasma*, *Plasmodium*, Apicomplexa, protein trafficking,
22 apicoplast, complex plastid.

23
24
25

26 ## Abstract
27
28 **Apicomplexa are unicellular parasites**
29 **causing important human and animal**
30 **diseases, including malaria and**
31 **toxoplasmosis. Most of these pathogens**
32 **possess a relict but essential plastid, the**
33 **apicoplast. The apicoplast was acquired by**
34 **secondary endosymbiosis between a red**
35 **alga and a flagellated eukaryotic protist. As**
36 **a result the apicoplast is surrounded by**
37 **four membranes. This complex structure**
38 **necessitates a system of transport signals**
39 **and translocons allowing nuclear encoded**
40 **proteins to find their way to specific**
41 **apicoplast sub-compartments. Previous**
42 **studies identified translocons traversing**
43 **two of the four apicoplast membranes. Here**
44 **we provide functional support for the role of**
45 **an apicomplexan Toc75 homolog in**
46 **apicoplast protein transport. We identify**
47 **two apicomplexan genes encoding Toc75**
48 **and Sam50, both members of the Omp85**
49 **protein superfamily. We localize the**
50 **respective proteins to the apicoplast and**
51 **the mitochondrion of *Toxoplasma* and**
52 **_Plasmodium_. We show that the *Toxoplasma***
53 **Toc75 is essential for parasite growth and**
54 **that its depletion results in a rapid defect in**
55 **the import of apicoplast stromal proteins**
56 **while the import of proteins of the outer**
57 **compartments is affected only as the**
58 **secondary consequence of organelle loss.**
59 **These observations along with the**
60 **homology of the protein to chloroplast**
61 **Toc75 suggest a role in transport through**
62 **the second innermost membrane.**
63
64

65 ## Introduction
66
67 Apicomplexan parasites are the cause of
68 important human and animal diseases,
69 including malaria and toxoplasmosis. Most of
70 these pathogens possess a relict plastid named
71 the apicoplast. While the apicoplast is no
72 longer photosynthetic, it has important
73 metabolic roles and supplies the parasite with
74 fatty acids, isoprenoids, and heme (1, 2). The
75 apicoplast is the product of secondary
76 endosymbiosis whereby a single celled red
77 alga was engulfed by a flagellated eukaryote
78 and a stable endosymbiotic relation ensued.
79 This event gave rise to a large and diverse
80 group of photosynthetic and non-photosynthetic
81 eukaryotes referred to by some authors as the
82 chromalveolates (3, 4). The apicoplast and the
83 plastids of other chromalveolates are
84 surrounded by four membranes reflecting their
85 complex endosymbiotic origin. The innermost
86 membrane and second innermost membrane
87 (SIMM) originate from the algal primary plastid.
88 The next membrane out, bounding the
89 periplastid compartment, originates from the
90 algal plasma-membrane and the outermost
91 membrane is believed to be derived from the
92 host endomembrane system (reviewed in (5)).
93 Key to the conversion of the algal
94 endosymbiont into a plastid was the transfer of
95 the symbiont's genes to the nucleus of the
96 host, allowing far reaching transcriptional and
97 translational control by the host. This transfer of
98 genetic material from the endosymbiont to the
99 host is only possible upon coevolving systems
100 that allow the import of host-translated proteins
101 into the endosymbiont. In the case of the
102 apicoplast this requires translocation across its
103 four delineating membranes to reach the

104 stroma. Apicomplexan parasites target large
105 numbers of nuclear encoded proteins to the
106 apicoplast. 10% of the *Plasmodium falciparum*
107 proteome is predicted to be transported to the
108 apicoplast, underscoring the importance of this
109 trafficking pathway (6). Our current model
110 (Figure 1A) assumes this pathway to start with
111 signal sequence guided entry into the ER
112 lumen, likely via the Sec61 translocon.
113 Trafficking from the ER to and across the outer
114 apicoplast membrane remains poorly
115 understood, but potentially depends on signals
116 typically involved in endocytosis or autophagy
117 (7-9) and may take place by more than one
118 route (10, 11). Translocation across the
119 periplastid membrane is mediated by
120 machinery evolved from the endosymbiont's
121 ER-associated protein degradation (ERAD)
122 system (12-16). Finally, based on their
123 evolutionary origin in chloroplast membranes, it
124 is believed that homologs of the translocons of
125 the inner and outer chloroplast membrane
126 (TIC/TOC) function in translocation of proteins
127 through the apicoplast's two innermost
128 membranes. Experimental evidence supports
129 the role of the TIC complex in apicoplast
130 protein import (17, 18) but is lacking in the case
131 of the putative TOC machinery.
132
133 Most stromal proteins possess a bipartite
134 signal, comprised of a signal and a transit
135 peptide (6). Upon translocation to the ER
136 lumen the N-terminal signal peptide portion of
137 the leader is cleaved off, exposing the transit
138 peptide that is required for further trafficking
139 (19). In diatoms, a group likely descended from
140 the same endosymbiotic event as
141 Apicomplexa, subplastidal targeting depends
142 on the first amino acid (position +1) of the
143 transit peptide (20, 21). An aromatic amino acid
144 at this position targets the protein through the
145 SIMM *en route* to the stroma; otherwise, the
146 proteins are retained in the periplastid space.
147 Incidentally, an aromatic residue is also
148 required for import through the outer
149 membrane of primary plastids of red algae (20,
150 22, 23) and of glaucocystophytes (24). In the
151 primary plastids of glaucocystophytes,
152 recognition of the aromatic residue depends on
153 an Omp85 family protein that functions as the
154 translocation pore of their primitive TOC
155 machinery (25).

156 Abundance of aromatic residues at position +1
157 was reported for the transit peptides of
158 additional Chromalveolates (26, 27). These
159 studies include *Toxoplasma* and *Plasmodium*
160 spp where enrichment of phenylalanine was
161 reported at this position (27). Nevertheless the
162 role of this amino acid was so far not supported
163 experimentally. The targeting sequence of a
164 *Toxoplasma* apicoplast stromal protein,
165 ferredoxin-NADP+ reductase *(*FNR), was
166 studied in detail (28). An extensive series of
167 deletions within the N-terminal sequence
168 suggested the presence of redundant signals
169 and did not implicate a particular residue at

170 position +1 (28). Whether this is true for all
171 apicoplast stromal proteins remains unknown.
172
173 Omp85 (for outer membrane protein of 85 KDa)
174 is a protein that catalyzes insertion and
175 assembly of β-barrel proteins into the outer
176 membrane of gram-negative bacteria. The
177 more widely distributed superfamily of Omp85-
178 related proteins shares a conserved domain
179 organization that includes N-terminal
180 polypeptide-transport-associated (POTRA)
181 repeats and a C-terminal transmembrane β-
182 barrel. Three main eukaryotic representatives
183 are well described: the mitochondrial sorting
184 and assembly machinery of 50 kDa
185 (Sam50/Tob55) and the chloroplast proteins
186 Toc75-III and Toc75-V (29, 30). Like its
187 bacterial ancestor, Sam50/Tob55 recognizes
188 mitochondrial outer membrane proteins in the
189 intermembrane space after they fully
190 translocate across the outer membrane and
191 catalyzes their insertion into it (31, 32). Toc75V
192 (or Oep80 for outer envelope protein 80) is
193 hypothesized to perform a similar role in the
194 outer chloroplast membrane (33-35). Toc75III
195 functions as the channel of the TOC translocon
196 in the outer chloroplast membrane that allows
197 proteins to fully translocate through it (36).
198 In diatoms, an Omp85-like protein, *Pt*Omp85,
199 was identified that possesses a bipartite plastid
200 targeting signal and two POTRA domains (37).
201 This protein is localized to the diatom complex
202 plastid and both its N and C terminal domain
203 face the periplastid compartment (37). Using
204 the sequence of *Pt*Omp85, Bullmann and co-
205 workers were able to identify putative
206 apicomplexan homologs (37, 38), and this
207 assignment gained further support from
208 Hirakawa and coworkers (39). These homologs
209 possess features supporting their Omp85
210 affiliation such as a signal sequence, the typical
211 N-terminal POTRA signature, and a C-terminus
212 that likely forms a beta-barrel. However, their
213 putative role in apicoplast protein import has
214 not been evaluated experimentally.
215 Here we seek to gain new insights into the
216 pathways by which apicoplast proteins traverse
217 the SIMM. We analyze the targeting sequences
218 of a large group of experimentally confirmed
219 apicoplast proteins (summarized in (40)), to
220 assess the abundance of an aromatic residue
221 that may be recognized by an Omp85. We
222 confirm the identity and localization of Omp85
223 proteins from both *Plasmodium falciparum* and
224 *Toxoplasma gondii* and demonstrate that the
225 *Toxoplasma* Toc75 functions in the import of
226 proteins into the stroma of the apicoplast.
227 Finally, we show that import of peripheral
228 apicoplast protein is not dependent on
229 TgToc75 activity, which is consistent with the
230 potential assignment of TgToc75 to the second
231 innermost of the four apicoplast membranes.

232
233 **Results**
234
235 *Sequence analysis reveals moderate
236 enrichment of aromatic residues at position +1*

of stromal proteins and the presence of two omp85-like proteins in apicomplexan genomes

We used sequence analysis to identify signals and machinery potentially involved in traversal of the apicoplast SIMM membrane. We have recently substantially expanded the repertoire of experimentally confirmed apicoplast proteins in *Toxoplasma gondii* (40) now counting 47 proteins (Table S1). We utilized the online prediction algorithm SignalP (http://www.cbs.dtu.dk/services/SignalP-3.0/) to predict the signal peptide cleavage site of all 47 proteins. Using SignalP 3.0 server we were able to define with high certainty the amino acid at position +1 of the transit peptide of 29 of these proteins (Table S1 shows the predictions obtained with both SignalP servers: 3.0 and 4.1). Figure 1B shows the distribution of +1 residue abundance in (i) 22 stromal and (ii) 7 peripheral proteins. We found that 27% of stromal proteins have an aromatic residue (mostly a phenylalanine) at the predicted position +1 while none of the non-stromal proteins feature an aromatic amino acid at this position (Table S1, Figure 1B).

Next, we revisited the repertoire of potential apicomplexan Omp85-like encoding genes. Using jackhmmer to mine the NCBI non-redundant database, and subsequent reciprocal BLAST searches against the EupathDB, we identified two Omp85-like proteins in *T. gondii* (TGME49_205570, TGME49_272390), *P. falciparum* (PF3D7_0608310, PF3D7_1234600), and several other apicomplexan species (Table 1). To determine the respective affiliation of these genes, we selected representative species across the eukaryotic tree of life and reconstructed a majority rule consensus tree from 1,000 bootstrap trees (Figure 2; see also maximum likelihood tree Figure S1A). The tree shows a clear split into Sam50 and Toc75 clades supported by a bootstrap of 100. We classified sequences TGME49_205570 and PF3D7_0608310 as Sam50 (herein named TgSam50 and PfSam50, respectively) and sequence TGME49_272390 as Toc75 (named TgToc75). PF3D7_1234600 could not be resolved with certainty in this analysis, and thus was not included in the reconstruction of this tree, however subsequent analysis included PF3D7_1234600 (Figure S1B) and used the POTRA region only (Figure S1C) to construct a maximum likelihood tree which shows that PF3D7_1234600 is affiliated with the Toc75 homologs from Chromalveolates (herein named PfToc75). The presence of two POTRA domains in PfToc75 and TgToc75 and a predicted apicoplast targeting signal in PfToc75 support this affiliation.

Mutagenesis of a phenylalanine at position +1 of the transit peptide of the stromal protein ACP to alanine results in peripheral retention

The putative role of the aromatic residue at position +1 of the stromal protein ACP in trafficking was analyzed via mutagenesis. YFP-tagged ACP with the wild type sequence (ACP$_{WT}$YFP) and YFP-tagged ACP with the phenylalanine replaced to an alanine (ACP$_{F/A}$YFP) were transiently transfected and their localization was assessed by high-resolution microscopy. While ACP$_{WT}$YFP showed precise co-localization with the stromal marker CPN60 (12), ACP$_{F/A}$YFP showed very little overlap with it (Figure 1C). Similarly, the signal from ACP$_{WT}$YFP did not overlap with signal from the HA-tagged periplastid marker ATrx2 (40), while ACP$_{F/A}$YFP showed substantial co-localization with this periplastid marker (Figure 1C).

The signal peptide cleavage prediction by SignalP 3.0 differs from that obtained by SignalP 4.1. While both suggest the phenylalanine at position +1 with high likelihood, the latter predicts an upstream tyrosine to be at this position. We generated YFP-tagged ACP with the tyrosine replaced to an alanine (ACP$_{Y/A}$YFP) and examined its localization upon transient expression by high-resolution microscopy. Similar to ACP$_{WT}$YFP, ACP$_{Y/A}$YFP showed full co-localization with the stromal marker CPN60 and little overlap with the periplastid marker ATrx2 (Figure 1C).

Localization of the T. gondii and P. falciparum Omp85 proteins to the apicoplast and mitochondrion supports their assignments as Toc75 and Sam50

The assignment of Omp85 proteins to their respective families as determined by the phylogeny was tested by localization studies. Both the first 78, and the first 95 N-terminal amino acids derived from both TgToc75 (Figure S2) and PfToc75 (Figure 3A) target to a punctate structure within the parasite that co-localized with the *Toxoplasma* or *Plasmodium* apicoplast markers FNR-RFP or ACP respectively. We conclude that these N-termini serve in apicoplast localization of these proteins. Moreover, full-length TgToc75 also co-localizes with the apicoplast marker FNR-RFP further supporting the Toc75 affiliation (Figure 3A). High resolution microscopy and co-staining with the stromal marker CPN60 suggested TgToc75 localization is peripheral to the apicoplast stroma (Figure 3B). In line with the expected peripheral localization of a Toc75 homolog.

We next assessed the localization of the second Omp85 homologue identified in each of the species. A mitochondrial targeting signal was predicted for PfSam50 but not for TgSam50 (Table S1). The first 60 amino acids of PfSam50 targeted GFP to a ribbon-like structure within *P. falciparum* parasites that co-localized with the signal obtained through the use of MitoTracker (Figure 3C). Likewise, full-length TgSam50 co-localized with the mitochondrial marker Hsp60-RFP (Figure 3C).

371 Finally, co-transfection of both full-length HA-
372 tagged TgToc75 and full-length Ty-tagged
373 TgSam50 in *T. gondii* reveals two distinct
374 patterns of fluorescence with minimal signal
375 overlap. This demonstrates the existence of
376 two Omp85-like proteins in *T. gondii* in the two
377 distinct endosymbiotic compartments; the
378 apicoplast and the mitochondrion (Figure 3D).
379 Taken together, our localization experiments
380 are entirely consistent with the classification
381 proposed by phylogenetic analysis.
382
383 *TgToc75 is required for parasite growth and*
384 *apicoplast maintenance*
385
386 To test whether TgToc75 functions in
387 apicoplast protein import we generated a
388 mutant in which its expression level can be
389 manipulated. First we constructed the
390 *TATiΔKu80iToc75pi* line: in this parasite the
391 TgToc75 open reading frame is separated from
392 its native promoter by a tetracycline-regulatable
393 promoter cassette (40). This parasite line was
394 established using cosmid (PSBL491)
395 recombineering (41, 42) (Figure S3). Our
396 analysis of this line suggested that TgToc75 is
397 essential for parasite growth (Figure S3), and
398 that its down regulation results in apicoplast
399 demise and in a stromal protein modification
400 defect (Figure S3), as expected from
401 interference in the apicoplast protein import
402 machinery (12, 18, 40). However, this mutant
403 proved unstable resulting in loss of regulation.
404 We thus utilized recombineering to construct
405 the *TATiΔKu80iToc75pr* line: in this line we
406 replaced the TgToc75 native promoter with the
407 tetracycline-regulatable promoter cassette
408 (Figure 4A). This line is stable and was used
409 for the remainder of the analyses. We found
410 down regulation of TgToc75 (Figure 4B) to
411 result in a growth defect as observed by plaque
412 assay (Figure 4C). Additionally, TgToc75
413 depletion resulted in loss of the apicoplast
414 evident by loss of plastid DNA, which was
415 quantified via quantitative PCR, as well as by
416 loss of immunofluorescence staining of the
417 apicoplast stromal protein CPN60 (Figure 4D,
418 E). Organelle loss was gradual starting with
419 28% loss at 24 hours of Toc75 down regulation
420 and reaching 99.5% loss by 96 hours.
421
422 *Loss of TgToc75 results in loss of import with*
423 *more rapid impact on stromal when compared*
424 *to peripheral apicoplast proteins*
425
426 To examine apicoplast protein import under
427 TgToc75 down regulation we followed the
428 maturation of the plastid stromal protein ACP
429 (12, 17, 18, 40). Typically two bands can be
430 observed for this protein by Western blot, a
431 larger precursor protein *en route* to the plastid,
432 and a mature protein lacking the leader peptide
433 due to the activity of stromal signal peptidase
434 (18, 19, 43, 44). By following endogenously
435 tagged ACP (40) we detected accumulation of
436 un-cleaved precursor starting at 48 hours after
437 Toc75 down regulation (Figure 5A).

438 Interestingly, the precursor of the protein
439 encoded by TGME49_001270, an outer
440 apicoplast membrane protein (40), does not
441 accumulate even as late as 72 hours (Figure
442 5B). To assess whether this difference is
443 specific to TgToc75 depletion we conducted
444 control experiments with a regulated mutant of
445 the periplastid protein 1 (PPP1). PPP1 is a
446 periplastid compartment resident protein that
447 plays an essential role in apicoplast protein
448 import and it is likely required for the
449 translocation of proteins across the periplastid
450 membrane (40). Here we show that upon down
451 regulation of PPP1 both the stromal ACP and
452 the outer membrane protein encoded by
453 TGME49_001270 show precursor
454 accumulation ((40) and Figure 5C,D). We
455 conclude that proteins pass through the PPP1
456 associated translocon first and the Toc75
457 translocon second and that the outer
458 translocon can act and assemble (at least for a
459 limited time) independently of Toc75.
460 To test whether these observations hold true
461 for other stromal and non-stromal proteins we
462 examined two additional markers, the stromal
463 protein LytB (45) and the periplastid protein
464 PPP1. In order to follow protein maturation in
465 real time we measured maturation of LytB and
466 PPP1 expressed transiently at different time
467 points after TgToc75 down regulation. In
468 agreement with the above observations, newly
469 synthesized stromal LytB, shows precursor
470 accumulation starting as early as 24 hours after
471 TgToc75 down regulation (Figure 5E), while the
472 newly synthesized periplastid protein PPP1
473 shows precursor accumulation only late into
474 suppression (72 hours, Figure 5F) when many
475 apicoplasts are lost due to secondary effects
476 (Figure 4D, E).
477
478 **Discussion**
479
480 The acquisition of secondary plastids went
481 hand in hand with the development of
482 appropriate machineries for protein import (3).
483 The complex nature of these plastids requires a
484 set of signals allowing precursor protein
485 trafficking to their final sub-organellar
486 destination. An elevated abundance of
487 aromatic amino acids, particularly
488 phenylalanine, at position +1 downstream of
489 the predicted signal peptide cleavage site, was
490 reported in several chromalveolates and was
491 proposed to be a functional feature of the
492 transit peptide in these organisms (26, 27). An
493 aromatic signature residue, most frequently a
494 phenylalanine (but also tyrosine and
495 tryptophan), at position +1 of the transit
496 peptide, was proposed to serve as forward
497 signaling from the periplastid space through the
498 two innermost membranes in several groups or
499 organisms with secondary plastids. A similar
500 requirement is found for import into the primary
501 plastids of red algae (20, 22, 23, 46). Gould
502 and coworkers suggested a model according to
503 which all proteins targeted to a complex plastid
504 of red origin gain entry to the periplastid

compartment by a common indiscriminate mechanism (46). We analyzed 47 *T. gondii* sequences of proteins experimentally shown to target to the apicoplast periphery or the apicoplast stroma. Of those we could assess 29 proteins with high certainty. This analysis does not align with the notion of a uniform mechanism. On one hand we show enrichment of an aromatic residue at position +1 in the putative transit peptides of proteins that cross all 4 apicoplast membranes (Table S1, Figure 1B). Further, our mutagenesis experiments support the idea that this aromatic +1 residue plays a role in the targeting of the stromal ACP (Figure 1C). However, on the other hand, not all stromal proteins obey this rule. In fact, the majority (73%) of stromal proteins were not predicted to have a +1 aromatic residue, suggesting alternative signals may be involved in SIMM traversal. Indeed, in the case of FNR, for which most computational analyzes ((28) + TableS1) do not predict an aromatic residue at position +1, other signals were implicated in stromal localization (28). We also observed the lack of aromatic residues at position +1 of peripheral proteins, however the repertoire of well documented residents of these outer compartments is still limited (only 7 predicted with confidence). Overall our observations are consistent with the previously proposed (20,21,27) broader conservation of the +1 aromatic signal as one of the mechanisms for stromal import but also suggest alternative, yet to be characterized, signals in Apicomplexa.

The secondary plastid of Apicomplexa and related taxa was shaped by contributions from three organisms: a cyanobacterium, a red alga and a flagellated heterotrophic eukaryote. The current model of protein import suggests that each membrane is traversed with the help of machinery derived from its organism of origin. This model gradually gained support with the identification and functional characterization of TIC components (17, 18) and of ERAD/SELMA components (12, 13, 16). The confirmation of the TOC link in this model was slow to emerge, most likely due to significant primary sequence divergence of the TOC components in organisms with complex plastids. An important breakthrough was made by the identification of an Omp85-like protein in the diatom *Phaeodactylum tricornutum,* for which phylogeny, subcellular localization and electrophysiology support affiliation with Toc75 (37). Here we provide experimental support for the general conservation of this transport pathway by localization of the apicomplexan homologs of PtToc75 to the apicoplast (Figure 3) and by functional analysis of TgToc75 (Figure 4 and 5).

Aside from TgToc75/PfToc75, our search for members of the polypeptide-transporting β-barrel protein superfamily in the genomes of Apicomplexa identified only one additional gene in each species, which encodes a Sam50/Tob55 homolog. We supported this assignment by localizing these proteins to the mitochondrion (Figure 3B). TgToc75/PfToc75 thus likely represent the only plastid Omp85s in these parasites, an observation that joins a growing line of evidence for a single Toc75 in the red lineage of plastids. The genome of the red alga *C. merolae* encodes a single Toc75 (47). Bullmann and coworkers similarly report a single Toc75 in their analysis of the genomes of the diatoms *P. tricornutum* and *Thalassiosira pseudonana,* and the haptophyte *Emiliania huxleyi* (37). In contrast, higher plants possess two functional Toc75 homologs: Toc75V/Oep80, which mediates assembly of proteins into the outer membrane of the chloroplast (35), and Toc75-III (36), which is the central component of the TOC machinery. At least two plastidial Toc75 proteins were identified in other members of the green lineage, and in all cases at least one ortholog of Toc75V/Oep85 was identified (39, 47). Whether the Toc75 found in the red lineage serves the roles of both of its green algal counterparts is unclear at this point.

Others (37) and we herein hypothesize that TgToc75 plays a role in precursor transit through the SIMM. To test TgToc75's involvement in apicoplast protein import we generated a conditional *TgToc75* mutant parasite cell using our recently described tetracycline-based promoter replacement system (40). We demonstrated *TgToc75* to be a firm requirement for apicoplast protein import, apicoplast maintenance, and parasite growth consistent with the hypothesis that this protein is an essential component of the apicoplast protein import machinery.

In agreement with a role for TgToc75 in stromal protein import, we observed a defect in precursor processing for the endogenously YFP-tagged stromal protein ACP (Figure 5A). The slow onset of this defect may reflect an overall slow impact of Toc75 mutants as previously noted in primary chloroplast (48), or could result from the long half-life of mature ACP as noted before (18, 40). We therefore tested an independent stromal protein (LytB) by transient transfection to follow the protein synthesized at various time points after Toc75 down regulation was ongoing. This assay showed impaired precursor processing as early as 24h after TgToc75 down regulation (Figure 5E) and before secondary defects due to loss of the organelle (Figure 4D,E). Overall the TgToc75 mutant produces a phenotype similar to previously studied inducible mutants in components of the apicoplast protein import machinery (12, 17, 18, 40) supporting the proposed role of TgToc75 in mediating stromal precursor protein import.

Interestingly, unlike the stromal proteins, only a mild processing defect was observed for outer compartment proteins (Figure 5B,F). This is specific to TgToc75 depletion, as processing is blocked for an outer compartment protein upon

638 disruption of the periplastid import machinery
639 (Figure 5D). These experiments support a
640 model under which proteins of the apicoplast
641 outer compartments (periplastid and outer
642 membrane compartments) are not dependent
643 on TgToc75 for their transport into the
644 organelle while stromal proteins (ACP and
645 LytB) are. Taken together with the phylogenetic
646 analyses these observations support TgToc75
647 as a component of the apicoplast TOC
648 channel, however direct experimental
649 demonstration for its activity at the SIMM is yet
650 to be established.
651 Finally, seeing that outer compartment protein
652 processing occurs under depletion of TgToc75,
653 our findings support the previous predictions
654 (15) of the existence of at least two apicoplast
655 transit peptide peptidases: one in the lumen
656 and one in the outer compartments upstream of
657 the TOC machinery.
658
659 While we provide functional support for the role
660 of Toc75 in protein import into complex plastids
661 of red origin, we were unable to identify other
662 components of the TOC machinery in the
663 genomes of Apicomplexa by using BLAST
664 searches, in line with previous reports (38, 47).
665 Most striking is the apparent absence of
666 homologs for the receptor components
667 Toc159/Toc120/Toc132 and Toc34/Toc33 (49)
668 that are found in primary plastids of both the
669 green and the red lineage (47, 50).
670 Interestingly, a similar finding was recently
671 reported for the secondary plastid of green
672 origin of *B. natans* (39). Hirakawa and
673 coworkers suggest an explanation whereby
674 unlike primary plastids where the TOC
675 machinery has to distinguish plastid proteins
676 from all other cytoplasmic and mitochondrial
677 proteins, the TOC machinery of secondary
678 plastids interacts with a more focused
679 repertoire of precursors that was already
680 screened by previous translocation
681 machineries. This idea is supported by the
682 observation that transit peptides of secondary
683 plastid apparently lack features that
684 differentiate between mitochondria and plastid
685 targeting in organisms with primary plastids
686 (27). In agreement with this model it was
687 proposed before that in membranes with a
688 primitive, reduced TOC machinery, the Omp85-
689 like component is involved in precursor
690 selection that is based on the presence of an
691 aromatic residue (25). While it is clear that
692 apicoplast stromal import could not be
693 explained by this simple model ((28), TableS1),
694 our finding provides grounds for further
695 investigation of the potential role of such a
696 pathway in the trafficking of at least some of
697 the stromal proteins.
698
699 One of the soluble components that interact
700 with the TOC machinery is Tic22 (51). TgTic22
701 was identified and functionally characterized
702 using a similar genetic system (17). TgTic22
703 down regulation results in a phenotype similar
704 to our observations here, whereby the

705 maturation of a stromal marker (FNR-DHFR-
706 cMyc) was reduced at 24h after addition of ATc
707 (17), supporting their potential cooperation in a
708 common pathway. Interestingly, Toc75 and
709 Tic22 are the sole TOC components found so
710 far in secondary red plastids. They are also the
711 only TOC components for which a clear
712 homology with their cyanobacterial ancestors
713 was demonstrated (51, 52).

Materials and methods

Search for Omp85 homologs
The non-redundant protein database was
downloaded from NCBI
(ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/) and
screened with jackhmmer (53) using AtToc75-
III as query for members of the Omp85
superfamily. Then *Toxoplasma gondii* ME49
and *Plasmodium falciparum* 3D7 genomes
were screened (i) with BLAST for homologs of
Toxoplasma and *Plasmodium* sequences
detected by jackhmmer and (ii) with
hmmsearch (53) for proteins with at least one
of the following PFAM (54) domains:
Surface_Ag_VNR (PF07244), Bac_surface_Ag
(PF01103), POTRA_2 (PF08479), ShlB
(PF03865). Finally the resulting *Plasmodium*
and *Toxoplasma* Omp85 homologs were used
as query for BLASTs of EupathDB
(http://eupathdb.org/eupathdb/) to identify the
respective homologs in other Apicomplexan
genomes.

Phylogenetic analysis
A multiple sequence consensus alignment was
constructed as described in (55) from a subset
of Sam50 and Toc75 homologs. From this
alignment a maximum likelihood phylogeny
was reconstructed with RAxML (56) using the
WAG model (57) and gamma-distributed rate
heterogeneity. Branch support values were
derived from 1,000 rapid bootstrap trees and a
majority rule consensus tree was constructed
from them. Note that the older gene model for
TgToc75, TGME49_072390, predicts a longer
protein, which includes the extreme C-terminal
part of the β-barrel (missing in the new gene
model: TGME49_272390). The alignment and
phylogeny were done with the longer older
gene model. Similarly, an older gene model of
PfSam50, PFF0410w, spans two new predicted
genes: PF3D7_0608300/0608310. We
experimentally confirmed the older gene model
(see supplementary text) new accession
numbers were produced for the confirmed
sequences (TgToc75 - KT271755; PfSam50 -
KT271756). The alignment and phylogeny were
performed using the new and shorter version
PF3D7_0608310, containing the conserved
domain.

Constructs:
Toxoplasma gondii:
Total RNA was extracted from *T. gondii* (strain
RH) using Trizol (Invitrogen). Overlapping
cDNA fragments encoding the entire TgToc75

and TgSam50 genes were amplified from total RNA using the SuperScript III One-Step RT-PCR kit (Invitrogen) and primers shown in Table S2. All resulting PCR products were cloned using the ZeroBlunt PCR cloning kit (Invitrogen) and sequenced (GATC, Konstanz). TgToc75[78], TgToc75[95]: Fragments encoding the noted amino acids were amplified from cDNA (primers in Table S2), and inserted into the TUB8mycGFPMyoATy *T. gondii* expression vector resulting in expression of these N-terminal amino acids directly fused to a Ty tag. Using EcoRI/NsiI allowed for an in frame C-terminal Ty tag. TgToc75[full-Ty]: A full-length cDNA version (based on the older gene model - TGME49_072390) of the *TgToc75* gene (removing an internal EcoRI restriction site) was synthesized by Geneart (Regensburg) and cloned into the TUB8mycGFPMyoATy vector as above. TgToc75[full-HA]: TgToc75[full-Ty] was digested with NsiI/PacI and a 3x hemagglutinin (HA) tag was inserted, having been generated by amplification (primers in Table S2). TgSam50[full-HA]: A full-length cDNA version of the *TgSam50* gene was synthesized by Geneart (Regensburg) and cloned into the TgToc75[full-HA] vector using EcoRI/NsiI.

Inducible knock-down cosmids:
pGDT7S4 (42) was used as templates to PCR amplify a 4Kb promoter modification cassette (primers in Table S2) containing a gentamycin resistance marker for selection in bacteria, a DHFR marker for the subsequent selection in *T. gondii* and the T7S4 promoter to be inserted upstream of TgToc75 start site (*pi*) or to replace the TgToc75 endogenous promoter sequence (*pr*). This cassette was used for PSBL491 recombineering as done before (41).

Site directed mutagenesis:
To change the residues at position +1 of the transit peptide of ACP within plasmid pTUB8ACP$_{WT}$YFP, primers ACP$_{F/A}$mutR or ACP$_{Y/A}$mutF/R (Table S2) were used in a site-directed mutagenesis reaction using the commercial QuikChange II Site-Directed Mutagenesis Kit (Stratagen) according to manufacturer's instructions.

Plasmodium falciparum:
Total RNA was extracted from *P. falciparum* (3D7) using Trizol (Invitrogen). PfToc75[78], PfToc[95], PfSam50[60]: Fragments encoding the noted amino acids were amplified from total RNA (primers in Table S2), and inserted into the XhoI/AvrII sites of pARL2-GFP. All final constructs were verified by restriction digest and automated sequencing (GATC, Konstanz).

Cell culture and transfection of T. gondii *and* P. falciparum.
Cultivation and transfections of *T. gondii* (strain RH delta hxgprt, a kind gift of Markus Meissner, and our TATi/ΔKu80strain (40)) in human foreskin fibroblasts and *P. falciparum* (3D7) in human erythrocytes was carried out under standard conditions. *P. falciparum* transfectants were selected with 2.5nM WR99210 (a kind gift of Jacobus Pharmaceuticals). Promoter replacement or insertion in TATi/ΔKu80strain was selected with 1µM Pyrimethamin as described in (42). FNR[RFP] and Hsp60[RFP] constructs were a kind gift of Markus Meissner.

Plaque assay
Fresh monolayers of HFF were infected with parasites in the presence or absence of 1.5 µg/ml ATc for 7 days. Fixation, staining and visualization were performed as previously described (40).

RT-PCR and qPCR
RNA was prepared from cultures grown without ATc or with ATc for 24 and 72 hours using RNeasy® (QIAGEN) and reverse transcriptase reaction was performed using SuperScript® III First-Strand Synthesis (Invitrogen) (both according to the manufacturer's instructions). 300ng of the resulting template was used for qPCR reaction using SYBR Green Mix (Bio-Rad) and primers TOC75RTPCRf2 and TOC75RTPCRr2. Copy number control was performed using cosmid PSBL491 as template. Genomic DNA was prepared from cultures grown without ATc or with ATc for 24 and 72 hours using DNeasy® (QIAGEN). 100ng of the resulting template was used for qPCR reaction using SYBR Green Mix (Bio-Rad) and primers Apg-qPCR-F/R for apicoplast and UPRT-qPCR-F/R for nuclear genomes. A copy number control was performed using specific plasmids as described in (58).

IFA and imaging
T. gondii: Immunofluoresence was carried out on infected HFF cells seeded onto glass cover slips. Cells were fixed with 4% paraformaldehyde/PBS (15 min, RT), permeabilised with 0.5% T-X-100/PBS (15 min, RT), blocked with 5% BSA/PBS (30 min, RT), incubated with primary antibodies diluted in 5% BSA/PBS (1h, RT), washed three times in PBS, incubated with suitable fluorescent-conjugated secondary antibodies (1h, RT), washed three times in PBS, incubated with 50 ng/ml Hoechst 33258/PBS (5 min, RT), washed in distilled water and cover slips were mounted onto glass slides using Fluoromount (SouthernBiotech).
P. falciparum: Cells were fixed in 4% Paraformaldehyde/0.00075% Glutaraldehyde (37ºC, 30 min), quenched in 125 mM Glycine/PBS, Hoechst 33258 (Molecular probes) was used at 50ng/ml for fixed parasites or 10 mg/ml for live parasites.
Images were acquired on Carl Zeiss Axio Observer inverse epifluorescence microscope (Figure3, FigureS2). Individual images were imported into ImageJ64 (version 1.46r, available at http://rsb.info.nih.gov/ij), converted to 8-bit greyscale, subjected to background subtraction, and overlaid. Image in FigureS3 was taken using a Delta Vision microscope as described (12). Antibodies and concentrations

used were: rabbit anti-HA (Sigma-Aldrich, 1:50); mouse anti-Ty tag (a kind gift of Keith Gull, 1:20); anti-ACP (a kind gift of Geoff McFadden, 1:500), rabbit anti-CPN60 (1:500), Cy2 goat anti-rabbit, Cy3 goat anti-Rabbit, Cy2 goat anti-mouse, Cy3 goat anti-mouse (all Jackson Immuno Research Laboratories, 1:2000).

For superresolution structural illumination microscopy (SR-SIM), stacks of 30-40 images were taken with increments of 0.091 µm in a Zeiss Elyra Superresolution microscope (Jena, Germany) with a 63x oil immersion objective and an immersion oil with a refractive index of 1.518 (Zeiss, Germany). Superresolution images were generated using ZEN software (version Zen 2012 SP1, Zeiss, Germany) and processed into their final form using FIJI software (59).

Apicoplast protein import assay and Western blot analyses:
Western blot of steady-state levels of proteins: clonal parasite lines grown in the presence or absence of ATc and collected (1500g, 10min, RT), lysed in sample buffer, separated by SDS-PAGE and blotted using anti-GFP (ROCHE) antibody for ACP-YFP and anti-HA antibody (SIGMA) for TGME49_001270.
Western blot of transiently expressed proteins: *TATiΔKu80iToc75pr* parasites were grown in ATc for a given period of time, then transiently transfected with pBT_LytB or pTUB8-PPP1-HA, and let to grow for an additional 24h to reach the total desired time of down-regulation (for example for 72 hours +ATc time point, parasites were grown for 48 hours in ATc, transfected and then grown for an additional 24h in ATc). Transfected and treated parasites were collected, separated by SDS-PAGE and blotted using anti-HA or anti-Ty antibodies.
Pulse/chase analysis was performed as described before (12, 18, 40).

References

1.   Sheiner L, Vaidya AB, McFadden GI. The metabolic roles of the endosymbiotic organelles of *Toxoplasma* and *Plasmodium* spp. Curr Opin Microbiol 2013;16(4):452-458.

2.   van Dooren GG, Striepen B. The algal past and parasite present of the apicoplast. Annu Rev Microbiol 2013;67:271-289.

3.   Cavalier-Smith T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol 1999;46(4):347-366.

4.   Koreny L, Sobotka R, Janouskovec J, Keeling PJ, Obornik M. Tetrapyrrole synthesis of photosynthetic chromerids is likely homologous to the unusual pathway of apicomplexan parasites. Plant Cell 2011;23(9):3454-3462.

5.   Sheiner L, Striepen B. Protein sorting in complex plastids. Biochim Biophys Acta 2013;1833(2):352-359.

6.   Foth BJ, Ralph SA, Tonkin CJ, Struck NS, Fraunholz M, Roos DS, Cowman AF, McFadden GI. Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. Science 2003;299(5607):705-708.

7.   Jayabalasingham B, Voss C, Ehrenman K, Romano JD, Smith ME, Fidock DA, Bosch J, Coppens I. Characterization of the ATG8-conjugation system in 2 *Plasmodium* species with special focus on the liver stage: possible linkage between the apicoplastic and autophagic systems? Autophagy 2014 ;10(2):269-284.

8.   Tawk L, Dubremetz JF, Montcourrier P, Chicanne G, Merezegue F, Richard V, Payrastre B, Meissner M, Vial HJ, Roy C, Wengelnik K, Lebrun M. Phosphatidylinositol 3-monophosphate is involved in *Toxoplasma* apicoplast biogenesis. PLoS Pathog 2011 ;7(2):e1001286.

9.   Tomlins AM, Ben-Rached F, Williams RA, Proto WR, Coppens I, Ruch U, Gilberger TW, Coombs GH, Mottram JC, Muller S, Langsley G. *Plasmodium falciparum* ATG8 implicated in both autophagy and apicoplast formation. Autophagy 2013;9(10):1540-1552.

10.   Bouchut A, Geiger JA, DeRocher AE, Parsons M. Vesicles bearing *Toxoplasma* apicoplast membrane proteins persist following loss of the relict plastid or Golgi body disruption. PLoS One 2014;9(11):e112096.

11.   Heiny SR, Pautz S, Recker M, Przyborski JM. Protein Traffic to the *Plasmodium falciparum* Apicoplast: Evidence for a Sorting Branch Point at the Golgi. Traffic 2014;15(12):1290-1304.

12.   Agrawal S, van Dooren GG, Beatty WL, Striepen B. Genetic evidence that an endosymbiont-derived endoplasmic reticulum-associated protein degradation (ERAD) system functions in import of apicoplast proteins. J Biol Chem 2009;284(48):33683-33691.

13.   Hempel F, Felsner G, Maier UG. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. Mol Microbiol 2010;76(3):793-801.

14.   Kalanon M, Tonkin CJ, McFadden GI. Characterization of two putative protein translocation components in the apicoplast of

Plasmodium falciparum. Eukaryot Cell 2009;8(8):1146-1154.

15. Spork S, Hiss JA, Mandel K, Sommer M, Kooij TW, Chu T, Schneider G, Maier UG, Przyborski JM. An unusual ERAD-like complex is targeted to the apicoplast of *Plasmodium falciparum*. Eukaryot Cell 2009;8(8):1134-1145.

16. Stork S, Moog D, Przyborski JM, Wilhelmi I, Zauner S, Maier UG. Distribution of the SELMA translocon in secondary plastids of red algal origin and predicted uncoupling of ubiquitin-dependent translocation from degradation. Eukaryot Cell 2012;11(12):1472-1481.

17. Glaser S, van Dooren GG, Agrawal S, Brooks CF, McFadden GI, Striepen B, Higgins MK. Tic22 is an essential chaperone required for protein import into the apicoplast. J Biol Chem 2012;287(47):39505-39512.

18. van Dooren GG, Tomova C, Agrawal S, Humbel BM, Striepen B. *Toxoplasma gondii* Tic20 is essential for apicoplast protein import. Proc Natl Acad Sci U S A 2008;105(36):13574-13579.

19. van Dooren GG, Su V, D'Ombrain MC, McFadden GI. Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. J Biol Chem 2002;277(26):23612-23619.

20. Gould SB, Sommer MS, Hadfi K, Zauner S, Kroth PG, Maier UG. Protein targeting into the complex plastid of cryptophytes. J Mol Evol 2006;62(6):674-681.

21. Gruber A, Vugrinec S, Hempel F, Gould SB, Maier UG, Kroth PG. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. Plant Mol Biol 2007;64(5):519-530.

22. Kilian O, Kroth PG. Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids. Plant J 2005;41(2):175-183.

23. Patron NJ, Waller RF, Archibald JM, Keeling PJ. Complex protein targeting to dinoflagellate plastids. J Mol Biol 2005;348(4):1015-1024.

24. Steiner JM, Yusa F, Pompe JA, Loffelhardt W. Homologous protein import machineries in chloroplasts and cyanelles. Plant J 2005;44(4):646-652.

25. Wunder T, Martin R, Loffelhardt W, Schleiff E, Steiner JM. The invariant phenylalanine of precursor proteins discloses the importance of Omp85 for protein translocation into cyanelles. BMC Evol Biol 2007;7:236.

26. Deane JA, Fraunholz M, Su V, Maier UG, Martin W, Durnford DG, McFadden GI. Evidence for nucleomorph to host nucleus gene transfer: light-harvesting complex proteins from cryptomonads and chlorarachniophytes. Protist 2000;151(3):239-252.

27. Ralph SA, Foth BJ, Hall N, McFadden GI. Evolutionary pressures on apicoplast transit peptides. Mol Biol Evol 2004;21(12):2183-2194.

28. Harb OS, Chatterjee B, Fraunholz MJ, Crawford MJ, Nishi M, Roos DS. Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite *Toxoplasma gondii*. Eukaryot Cell 2004;3(3):663-674.

29. Schleiff E, Becker T. Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat Rev Mol Cell Biol 2011;12(1):48-59.

30. Simmerman RF, Dave AM, Bruce BD. Structure and function of POTRA domains of Omp85/TPS superfamily. Int Rev Cell Mol Biol 2014;308:1-34.

31. Gentle I, Gabriel K, Beech P, Waller R, Lithgow T. The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. J Cell Biol 2004;164(1):19-24.

32. Kozjak V, Wiedemann N, Milenkovic D, Lohaus C, Meyer HE, Guiard B, Meisinger C, Pfanner N. An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. J Biol Chem 2003;278(49):48520-48523.

33. Eckart K, Eichacker L, Sohrt K, Schleiff E, Heins L, Soll J. A Toc75-like protein import channel is abundant in chloroplasts. EMBO Rep 2002;3(6):557-562.

34. Inoue K, Potter D. The chloroplastic protein translocation channel Toc75 and its paralog OEP80 represent two distinct protein families and are targeted to the chloroplastic outer envelope by different mechanisms. Plant J 2004;39(3):354-365.

35. Schleiff E, Soll J, Kuchler M, Kuhlbrandt W, Harrer R. Characterization of the translocon of the outer envelope of chloroplasts. J Cell Biol 2003;160(4):541-551.

36. Schnell DJ, Kessler F, Blobel G. Isolation of components of the chloroplast protein import machinery. Science 1994;266(5187):1007-1012.

37. Bullmann L, Haarmann R, Mirus O, Bredemeier R, Hempel F, Maier UG, Schleiff E. Filling the gap, evolutionarily conserved Omp85 in plastids of chromalveolates. J Biol Chem 2010;285(9):6848-6856.

38. Agrawal S, Striepen B. More membranes, more proteins: complex protein import mechanisms into secondary plastids. Protist 2010;161(5):672-687.

39. Hirakawa Y, Burki F, Keeling PJ. Genome-based reconstruction of the protein import machinery in the secondary plastid of a chlorarachniophyte alga. Eukaryot Cell 2012;11(3):324-333.

40. Sheiner L, Demerly JL, Poulsen N, Beatty WL, Lucas O, Behnke MS, White MW, Striepen B. A systematic screen to discover and analyze apicoplast proteins identifies a conserved and essential protein import factor. PLoS Pathog 2011;7(12):e1002392.

41. Brooks CF, Johnsen H, van Dooren GG, Muthalagi M, Lin SS, Bohne W, Fischer K,

Striepen B. The *Toxoplasma* apicoplast phosphate translocator links cytosolic and apicoplast metabolism and is essential for parasite survival. Cell Host Microbe 2010;7(1):62-73.

42. Francia ME, Jordan CN, Patel JD, Sheiner L, Demerly JL, Fellows JD, de Leon JC, Morrissette NS, Dubremetz JF, Striepen B. Cell division in Apicomplexan parasites is organized by a homolog of the striated rootlet fiber of algal flagella. PLoS Biol 2012;10(12):e1001444.

43. Waller RF, Keeling PJ, Donald RG, Striepen B, Handman E, Lang-Unnasch N, Cowman AF, Besra GS, Roos DS, McFadden GI. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc Natl Acad Sci U S A 1998;95(21):12352-12357.

44. Waller RF, Reed MB, Cowman AF, McFadden GI. Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. EMBO J 2000;19(8):1794-1802.

45. Nair SC, Brooks CF, Goodman CD, Sturm A, McFadden GI, Sundriyal S, Anglin JL, Song Y, Moreno SN, Striepen B. Apicoplast isoprenoid precursor synthesis and the molecular basis of fosmidomycin resistance in *Toxoplasma gondii*. J Exp Med 2011;208(7):1547-1559.

46. Gould SB, Sommer MS, Kroth PG, Gile GH, Keeling PJ, Maier UG. Nucleus-to-nucleus gene transfer and protein retargeting into a remnant cytoplasm of cryptophytes and diatoms. Mol Biol Evol 2006;23(12):2413-2422.

47. Kalanon M, McFadden GI. The chloroplast protein translocation complexes of *Chlamydomonas reinhardtii*: a bioinformatic comparison of Toc and Tic components in plants, green algae and red algae. Genetics 2008;179(1):95-112.

48. Huang W, Ling Q, Bedard J, Lilley K, Jarvis P. In vivo analyses of the roles of essential Omp85-related proteins in the chloroplast outer envelope membrane. Plant Physiol 2011;157(1):147-159.

49. Jarvis P, Chen LJ, Li H, Peto CA, Fankhauser C, Chory J. An Arabidopsis mutant defective in the plastid general protein import apparatus. Science 1998;282(5386):100-103.

50. Reumann S, Inoue K, Keegstra K. Evolution of the general protein import pathway of plastids (review). Mol Membr Biol 2005;22(1-2):73-86.

51. Tripp J, Hahn A, Koenig P, Flinner N, Bublak D, Brouwer EM, Ertel F, Mirus O, Sinning I, Tews I, Schleiff E. Structure and conservation of the periplasmic targeting factor Tic22 protein from plants and cyanobacteria. J Biol Chem 2012;287(29):24164-24173.

52. Bredemeier R, Schlegel T, Ertel F, Vojta A, Borissenko L, Bohnsack MT, Groll M, von Haeseler A, Schleiff E. Functional and phylogenetic properties of the pore-forming beta-barrel transporters of the Omp85 family. J Biol Chem 2007;282(3):1882-1890.

53. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol 2011;7(10):e1002195.

54. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. Pfam: the protein families database. Nucleic Acids Res 2014;42(Database issue):D222-230.

55. Flinner N, Ellenrieder L, Stiller SB, Becker T, Schleiff E, Mirus O. Mdm10 is an ancient eukaryotic porin co-occurring with the ERMES complex. Biochim Biophys Acta 2013;1833(12):3314-3325.

56. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30(9):1312-1313.

57. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 2001;18(5):691-699.

58. Reiff SB, Vaishnava S, Striepen B. The HU protein is important for apicoplast genome maintenance and inheritance in *Toxoplasma gondii*. Eukaryot Cell 2012;11(7):905-915.

59. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, *et al.* Fiji: an open-source platform for biological-image analysis. Nat Methods 2012;9(7):676-682.

60. Agrawal S, Chung DW, Ponts N, van Dooren GG, Prudhomme J, Brooks CF, Rodrigues EM, Tan JC, Ferdig MT, Striepen B, Le Roch KG. An apicoplast localized ubiquitylation system is required for the import of nuclear-encoded plastid proteins. PLoS Pathog 2013;9(6):e1003426.

61. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33(Web Server issue):W244-248.

Figure legends

Figure 1 *Machinery and signals involved in the translocation of precursor protein through the apicoplast membranes*. (A) Schematic representation of the translocation machinery responsible for protein import through the four membranes of the apicoplast. Endomembrane system shown in grey, former red algal cytosol in blue and former primary plastid in pink. According to the current model of transport apicoplast precursor proteins are first co-translationally transported into the ER via the SEC61 complex courtesy of their signal peptide (SP). In the ER lumen, the SP is cleaved by a signal peptide peptidase (SiPP). The now exposed transit peptide (TP) signals for transport from the ER to the apicoplast. Next, the translocation through ERAD/SELMA is likely accompanied by transient ubiquitination (60). The protein then moves through the TOC and TIC complexes and its TP is cleaved in the stroma. Compartmental markers used in this study are depicted at the upper right corner of their corresponding compartment. (B) Abundance of residues at position +1 of 29 proteins experimentally confirmed to localize to the apicoplast stroma (i) or peripheral compartments (ii) based on SP cleavage prediction by SignalP (detailed analysis is provided in TableS1). (C) High-resolution microscopy analysis of the localization of transiently expressed ACP-YFP with the wild type phenylalanine and tyrosine at the two predicted potential position +1 (i) with tyrosine to alanine mutation (ii) and with phenylalanine to alanine mutation (iii). Full SP sequence is shown with arrows showing both potential +1 residues in the wild type, or the position of mutation to alanine in the mutants. The upper panels show co-staining with the stromal marker CPN60 and the lower panels show co-staining with the PPC marker ATrx2 (40).

Figure 2 *Phylogenetic classification of Omp85-like proteins in* T. gondii *and* P. falciparum. A majority rule consensus tree of selected Sam50 and Toc75 homologs was constructed with RAxML from 1,000 bootstrap trees. The corresponding maximum likelihood (ML) tree is given in Figure S1A. The proteins are referenced by their ID (GenBank, or EupathDB); species abbreviations are as follows: Anig (*Aspergillus niger* ATCC 1015), Atha (*Arabidopsis thaliana*), Bden (*Batrachochytrium dendrobatidis* JAM81), Cmer (*Cyanidioschyzon merolae* strain 10D), Cowc (*Capsaspora owczarzaki* ATCC 30864), Crei (*Chlamydomonas reinhardtii*), Dmel (*Drosophila melanogaster*), Esil (*Ectocarpus siliculosus*), Gsul (*Galdieria sulphuraria*), Hvul (*Hydra vulgaris*), Mocc (*Metaseiulus occidentalis*), Otau (*Ostreococcus tauri*), Pfal (*Plasmodium falciparum* 3D7), Pmar (*Perkinsus marinus* ATCC 50983), Ppat (*Physcomitrella patens*), Ptri (*Phaeodactylum tricornutum* CCAP 1055/1), Rirr (*Rhizophagus irregularis* DAOM 197198w), Rnor (*Rattus norvegicus*), Scer (*Saccharomyces cerevisiae* S288c), Skow (*Saccoglossus kowalevskii*), Spur (*Strongylocentrotus purpuratus*), Tgon (*Toxoplasma gondii* ME49), Tura (*Triticum urartu*), Vcar (*Volvox carteri* f. nagariensis). The full alignments used for this analysis are provided in Figure S1.

Figure 3. *Localization of the omp85-like proteins supports their affiliation as Toc75 and Sam50 in Apicomplexa.* (A) Fluorescence microscopy analysis of *P. falciparum* parasites expressing ectopic GFP fusions of the 78 (upper panel) and 95 (middle panels) N-terminal amino acids of PfToc75, and of *T. gondii* parasites expressing ectopic HA-tagged full-length TgToc75 (lower panel). Co-staining is done with ACP and FNR-RFP for *P. falciparum* and *T. gondii* respectively. (B) High-resolution microscopy of transiently expressed full-length Ty tagged TgToc75 and its localization with respect to the stromal marker CPN60. (C) *P. falciparum* parasite expressing ectopic GFP fusion of the 60 N-terminal amino acids of PfSam50 (green channel) co-stained with mito-tracker (red channel) (upper panel); *T. gondii* parasites expressing ectopic Ty-tagged TgSam50 (green channel) co-stained with the mitochondrial marker HSP60-RFP (red channel) (lower panel). (D) *T. gondii* parasites co-expressing ectopic HA-tagged TgToc75 (red channel) and Ty-tagged TgSam50 (green channel).

Figure 4. *TgToc75 is essential for parasite growth and apicoplast maintenance*. (A) Schematic representation of the manipulation of the *TgToc75* locus to replace the native promoter with the tetracycline inducible promoter. Black boxes – exons; asterisk – stop codon; empty boxes – minigenes; solid lines – TgToc75 locus non-coding sequences; dashed line – genomic sequence; grey thick line – backbone of cosmid or of modification cassette. (B) Plaque assays performed with the *TATiΔKu80iToc75pr* parasite line in the absence (-) or presence (+) of ATc. (C) qRT-PCR analysis with RNA extracts from *TATiΔKu80iToc75pr* grown in the absence of ATc (-ATc) or upon down regulation of TgToc75 for 24 (+24h) and 72 (+72h) hours. TgToc75 mRNA levels decline swiftly upon ATc treatment. Y-axis shows the percentage of wild type copy numbers. (D) *TATiΔKu80iToc75pr* parasites were grown in ATc as indicated and plastids were counted based on immunofluorescence signal obtained via staining with anti-CPN60 antibody in 100 four-parasites vacuoles for each sample. Y-axis shows percentage of 4-parasites-vacuoles. (E) Apicoplast loss was also evaluated using qPCR comparing nuclear genome and apicoplast genome copy numbers. The data was normalized such that copy number from each genome from no ATc treatment is 1. In support of apicoplast loss the proportion of apicoplast copy number after TgToc75 down-regulation for 72 hours is on average 0.17 while genomic copy number average proportion is 0.78.

Figure 5. *TgToc75 down regulation results in deficient import of stromal but not PPC or outer membrane compartment apicoplast proteins*. We performed western blot analysis to follow the

maturation of apicoplast proteins under the down regulation of apicoplast import components. The steady state expression of endogenously YFP-tagged ACP (40) (A) and of endogenously HA-tagged TGME49_001270 (B) was monitored at each time point of *TgToc75* down regulation showing maturation defect in ACP but not TGME49_001270 at 72 hours. Western blot analysis following the maturation of the same makers (YFP-tagged ACP (C) and endogenously HA-tagged TGME49_001270 (D)), but this time under down regulation of the PPC import component *TgPPP1,* shows maturation defect for both at 48 hours. We then performed western blot analysis following the maturation of the stromal protein LytB-Ty (E) and the periplastid protein PPP1 (F). In this experiment LytB or PPP1 are transiently expressed for 24h at each time point of *TgToc75* down regulation. This analysis reveals a block in LytB maturation that is first detected at 24 hours and complete by 48 hours. In contrast, maturation defect of PPP1 is only observed at 72 hours. Loading control performed with anti-alpha-tubulin antibody is shown for each blot.

Figure S1 – *Phylogenetic classification of Omp85-like proteins in* T. gondii *and* P. falciparum*.*
Maximum likelihood (ML) phylogenies were reconstructed with RAxML. Branch support values were determined from 1,000 bootstrap trees. Sam50 and Toc75 clades are marked by dark and light gray areas, respectively. The sequence labels are colored according to their taxonomy (color code given in Figure 2). (A) The ML tree was reconstructed from the same set of sequences as used for the majority rule consensus tree in Figure 2. (B) This ML tree was reconstructed with the same sequences as (A) while including PfToc75. (C) From the multiple sequence consensus alignment as used for the trees above the N-terminal part containing the POTRA domains was excised and a ML tree reconstructed. (D) The full alignments used for these phylogenies (also available in other formats upon request).

Figure S2 – *The N-terminal domain of TgToc75 is sufficient for apicoplast localization.* Fluorescence microscopy analysis of parasites expressing ectopic Ty-tagged fusions of the 78 (upper panel) and 95 (lower panels) N-terminal amino acids of TgToc75. Note that the Ty tags are directly fused to the 78 or 95 amino acids with no spacer sequences.

Figure S3 – *TgToc75 is essential for parasite growth and apicoplast biogenesis*. (A) Schematic representation of the manipulation of the *TgToc75* locus to insert the tetracycline inducible promoter between the native promoter and the ORF. (B) Plaque assays performed with the *TATiΔKu80iToc75pi* parasite line in the absence (-) or presence (+) of ATc which correspond to TgToc75 constitutive levels or down-regulation respectively. (C) Fluorescence microscopy of *TATiΔKu80iToc75pi* grown in absence of ATc (-ATc) or upon down-regulation of TgToc75 for 72 hours (+ATc 72h) stained with the apicoplast marker CPN60 (12) showing loss of apicoplast in most parasites at this time point. (D) Pulse-chase (*P/C*) analysis of protein synthesis and post-translational lipoylation of apicoplast (PDH-E2) and mitochondrial (mito-E2) proteins. Parasites were metabolically labeled as detailed in (12, 18, 40) and lipoylated proteins were isolated by immunoprecipitation using a specific antibody. Lipoylation of PDH-E2 is lost upon ATc treatment. Bands labeled with an asterisk likely represent lipoylated host cell proteins

Table 1 – GeneIDs and summary of targeting prediction for apicomplexan Omp85-like protein encoding genes

Table S1 – Prediction of signal peptide cleavage and amino acid at position +1 of putative transit peptide for 47 experimentally confirmed apicoplast proteins.

Table S2 – Primers used in this study

1
2

List of supplemental materials:

Material included	Main text associate	Significance
Text + Figure S1	Result paragraph 1	Provides detailed explanation on sequence identification experimental confirmation and phylogenetic analysis allowing expert reader to critically follow the process. Provides the alignments used for the phylogenetic analyses.
Table S1	Figure 1B	Raw data of results summarized in the graphs. Reader can extract more information: the specific gene IDs used and the scores for each data point.
Table S2	Materials and methods	List of all primers used for genetic manipulations described in the text. Technical details for reader who wishes to perform similar manipulations.
Figure S2	Figure 3	Allows the reader to compare the localization pattern observed with the N-terminal fusion to the full-length that appears in the main text. In some organisms N-terminal fusion is more common. Showing that both generate the same localization validates this approach.
Figure S3	Figure 4	Provide evidence to an important difference between two approaches of genetic manipulation that are commonly used in *T. gondii*. Provides an additional independent assessment of Toc75's role in apicoplast biogenesis.

3
4 **Supplementary text**
5
6 *Sequence analysis of omp85-like proteins in Apicomplexa genomes*
7
8 We used jackhmmer to mine the NCBI non-redundant database with *Arabidopsis thaliana* Toc75-III as
9 query sequence. This search revealed two Omp85-like protein coding genes in *T. gondii* ME49
10 (TGME49_205570, TGME49_272390) and in several *Plasmodium* spp (Table 1). Subsequent Omp85-
11 related pHMM searches in the PFAM database, and a BLAST search using the above detected
12 *Toxoplasma* and *Plasmodium* Omp85 proteins unraveled two Omp85-like proteins also in *P. falciparum*
13 3D7 (PF3D7_0608310 and PF3D7_1234600). Reciprocal BLASTs against the apicomplexan databases
14 in EupathDB (http://eupathdb.org/eupathdb/) identified further homologs of both proteins encoded by
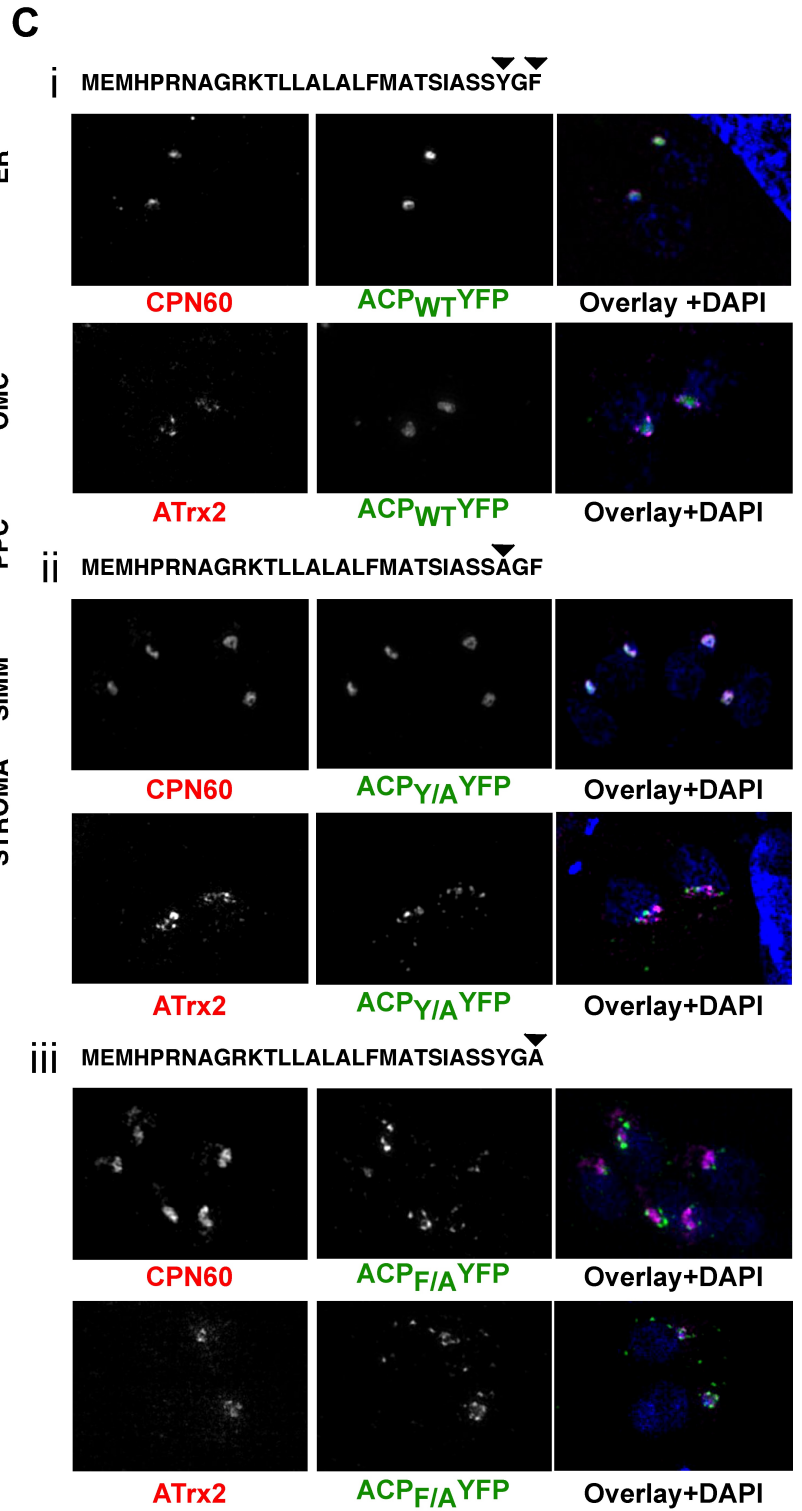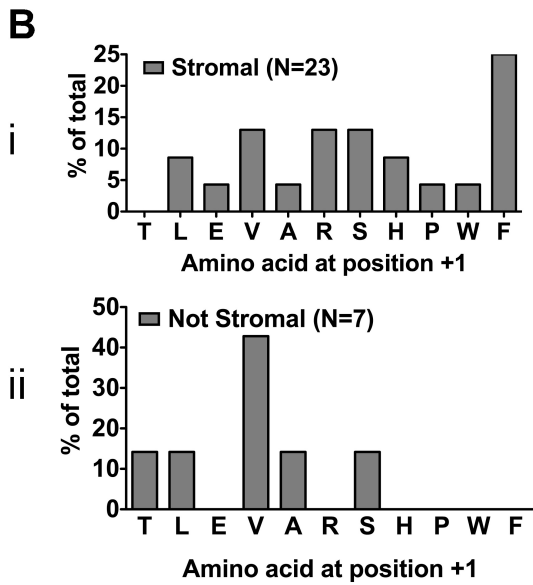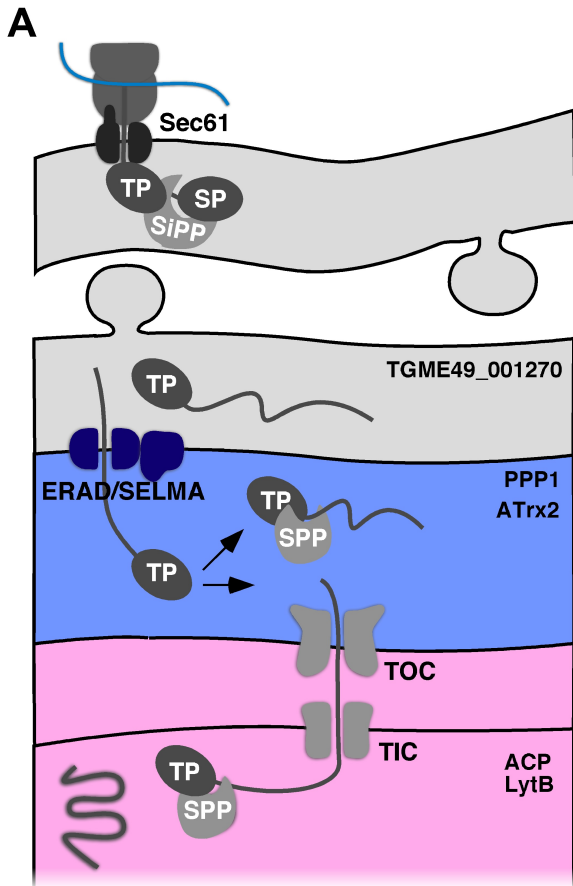15 several species (Table 1).
16
17 The predicted gene models for TgToc75 and PfToc75 as found on EupathDB were changed since we
18 first identified these genes: TgToc75 older version, TGME49_072390 includes an extreme C-terminal
19 domain, which is part of the predicted β-barrel. This C-terminal domain is missing in the new version
20 (TGME49_272390). Our RT-PCR and localization of full-length protein supports the old gene models.
21 Similarly, PfToc75's previous model (PFF0410w) predicts one continuous gene, which is now predicted
22 to be two separate genes (PF3D7_0608300/0608310). Our RT-PCR confirms the old model. Prediction
23 of organelle targeting signals as shown in table 1 used the older experimentally confirmed gene models.
24 User comments were added to the respective gene pages in ToxoDB and PlasmoDB.
25
26 To determine the affiliations of the four identified Omp85-like sequences from *Plasmodium falciparum*
27 and *Toxoplasma gondii*, we selected a subset of species across the eukaryotic tree of life and
28 generated a sequence alignment as described in (55). A majority rule consensus tree was then
29 constructed from 1,000 bootstrap trees based on this alignment (Figure 2). A maximum likelihood (ML)
30 phylogeny was reconstructed from the same dataset with RAxML. Branch support values were
31 determined from 1,000 bootstrap trees (Figure S1A). We then added the second *Plasmodium*
32 *falciparum* Omp85 sequence that was not originally identified via the jackhmmer search
33 (PF3D7_1234600) to the dataset and reconstructed another ML tree (Figure S1B). However, the
34 classification of this sequence is ambiguous. We set out to clarify its affiliation by constructing a
35 phylogenetic tree of the excised POTRA region, which is more conserved than the β-barrel region and
36 thus more suitable for the tree reconstruction. This tree shows that PF3D7_1234600 is located within
37 the sub-tree containing the other Toc75 homologs from Chromalveolates (Figure 2B). Furthermore, the
38 bootstrap between the Sam50 and Toc75 clades is reliable with a value of 88. In our alignment we could
39 identify two POTRA domains in PfToc75 (residues 118-192, 193-454) and TgToc75 (189-336, 337-451),
40 which are in agreement with fold recognition results except that the HHpred webserver (61) does not
41 detect the 1st β-strand of PfToc75's 2nd POTRA domain (399-454). In agreement with the assignment
42 based on the phylogenetic trees PfToc75 possesses a predicted apicoplast-targeting signal and for
43 PfSam50 a mitochondrial targeting sequence was predicted (Table 1).
44
45 **Supplementary references**

Sheiner et. al.

1   1.   Flinner N, Ellenrieder L, Stiller SB, Becker T, Schleiff E, Mirus O. Mdm10 is an ancient
2   eukaryotic porin co-occurring with the ERMES complex. Biochim Biophys Acta 2013;1833(12):3314-
3   3325.
4   2.   Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection
5   and structure prediction. Nucleic Acids Res 2005;33(Web Server issue):W244-248.
6

Figure 1

A

B

i

ii

C

i MEMHPRNAGRKTLLALALFMATSIASSYGF

CPN60 ACP_WT_YFP Overlay +DAPI

ATrx2 ACP_WT_YFP Overlay+DAPI

ii MEMHPRNAGRKTLLALALFMATSIASSAGF

CPN60 ACP_Y/A_YFP Overlay+DAPI

ATrx2 ACP_Y/A_YFP Overlay+DAPI

iii MEMHPRNAGRKTLLALALFMATSIASSYGA

CPN60 ACP_F/A_YFP Overlay+DAPI

ATrx2 ACP_F/A_YFP Overlay+DAPI

Figure 2

Figure 3

A

Hoechst	PfToc75 $^{1-78}$	αACP	Overlay	DIC Overlay
Hoechst	PfToc75 $^{1-95}$	αACP	Overlay	DIC Overlay
Hoechst	TgToc75-HA	FNR-RFP	Overlay	DIC Overlay

B

| Hoechst | TgToc75-Ty | CPN60 | Overlay | Inset |

C

| Hoechst | PfSam50^{N}GFP | MitoTracker | Overlay | DIC Overlay |
| Hoechst | TgSam50-Ty | HSP60-RFP | Overlay | DIC Overlay |

D

| Hoechst | TgToc75-HA | TgSam50-Ty | Overlay | DIC Overlay |

Figure 4

Figure 5

Steady state protein level

A

ACP (stromal)/ΔToc

-ATc +24 +48 +72

pre - m —
mature —
41KDa

Tub control

B

001270 (OMM)/ΔToc

-ATc +24 +48 +72

pre - m —
mature —
98KDa
62KDa

Tub control

C

ACP (stromal)/ΔPPP1

pre - m —
mature —
41KDa

Tub control

D

001270 (OMM)/ΔPPP1

pre - m —
mature —
98KDa
62KDa

Tub control
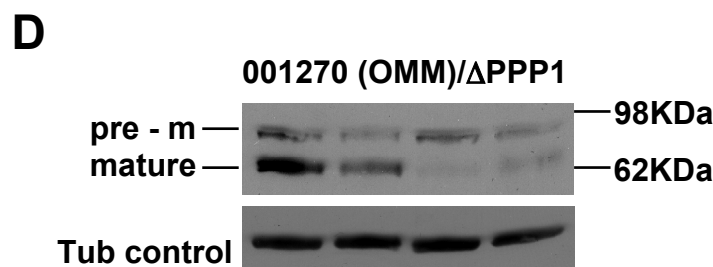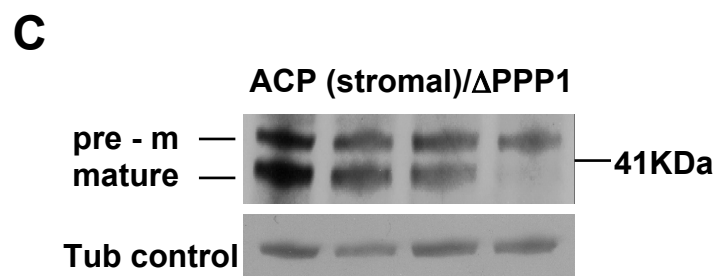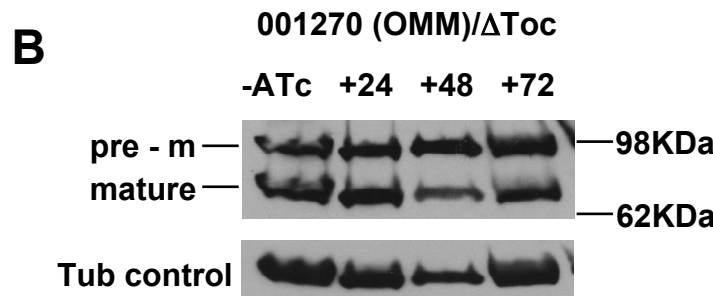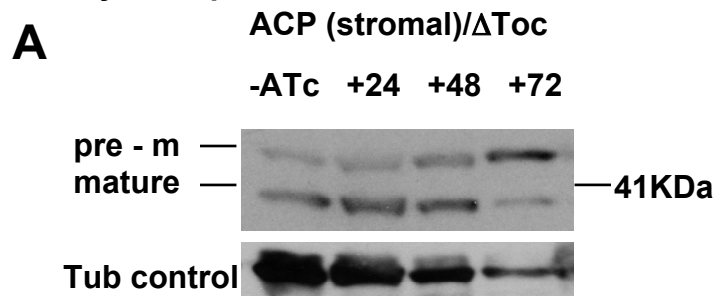
Transient expression at real time

E

LytB (stromal)/ΔToc

pre - m —
mature —
98KDa
62KDa

Tub control

F

PPP1 (PPC)/ΔToc

pre - m —
mature —
62KDa
41KDa

Tub control
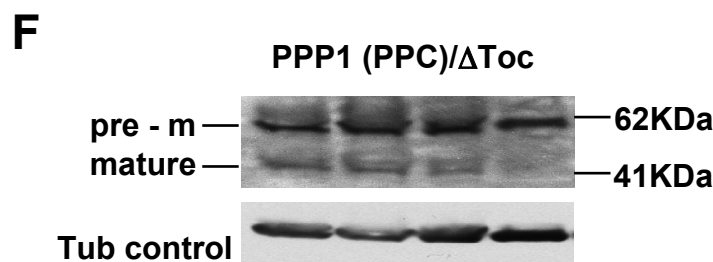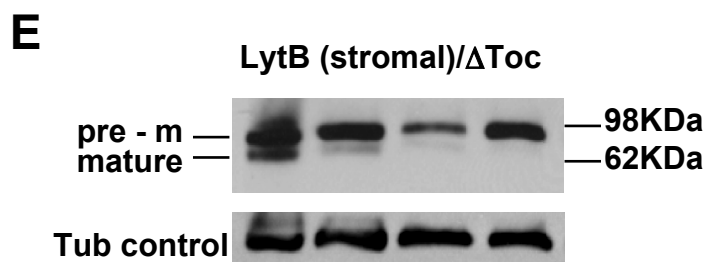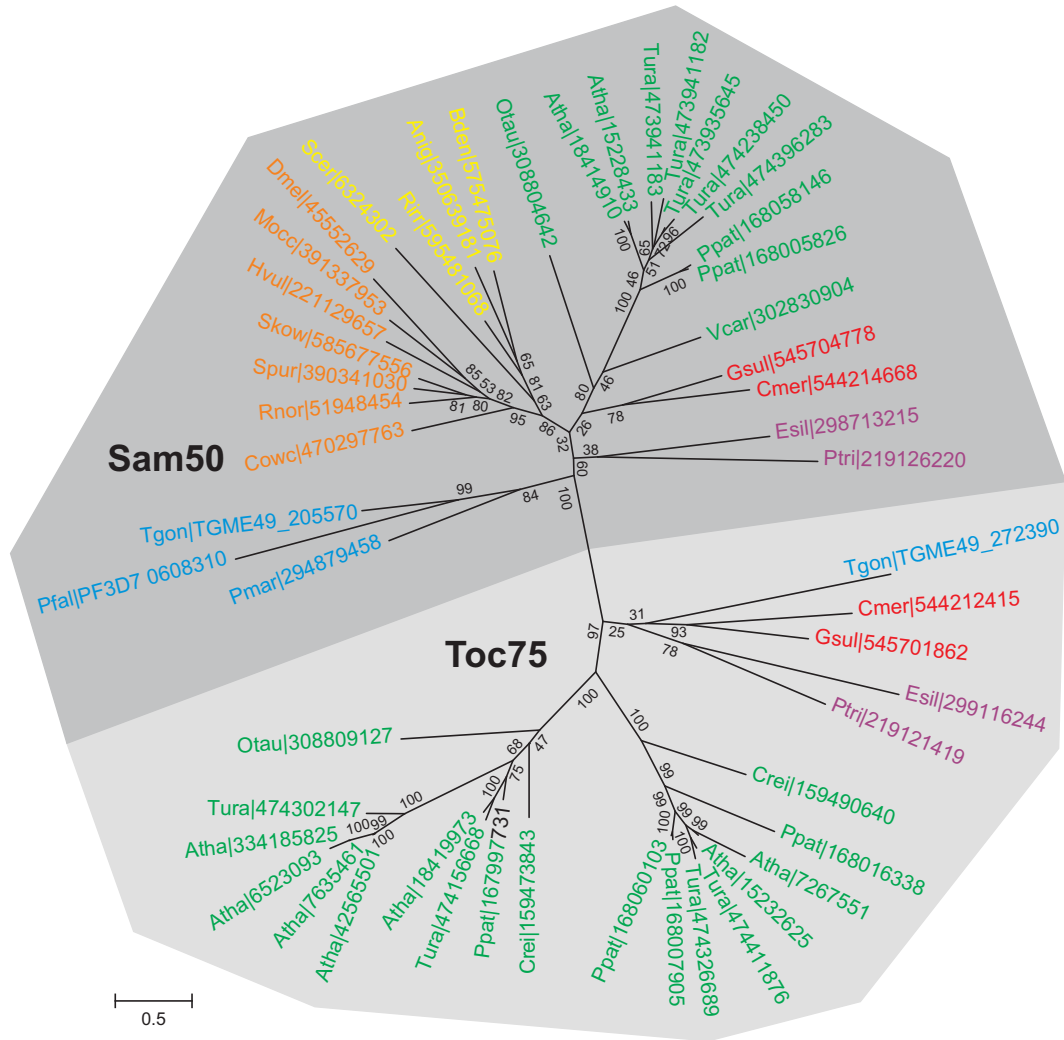
Figure S1A

Figure S1B

Figure S1C

Figure S2

| Hoechst | TgToc75 1-78 | FNR-RFP | DIC Overlay | Overlay |

| Hoechst | TgToc75 1-95 | FNR-RFP | DIC Overlay | Overlay |

Figure S3

A

modified cosmid

Genomic locus

Modified locus

Key

T7S4 — Inducible promoter

TgToc75 promoter

TgToc75 gene

B

-ATc

+ATc

C

- ATC: DIC | DAPI | CPN60

+ ATc 72h: DIC | DAPI | CPN60

D

- ATC   + ATc 72h
P   C   P   C

*

PDH-E2

*

Mito-E2

Mito-E2
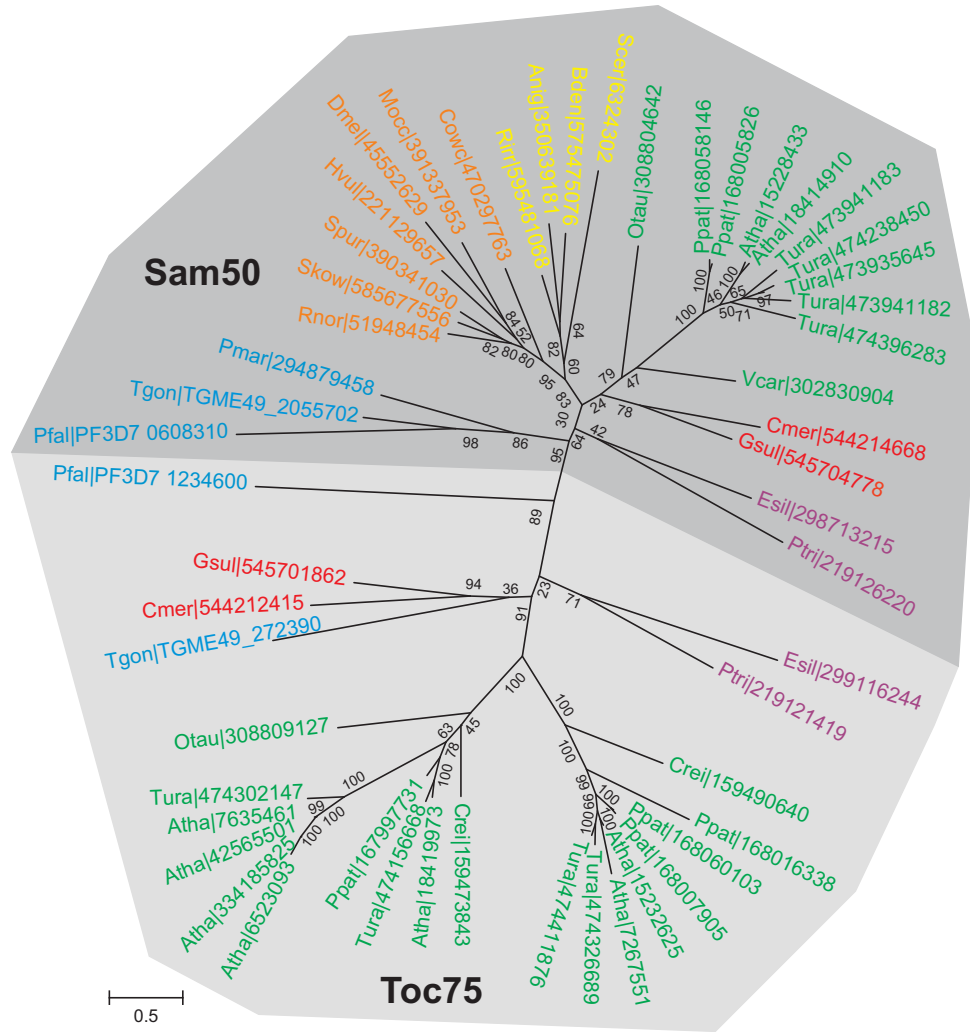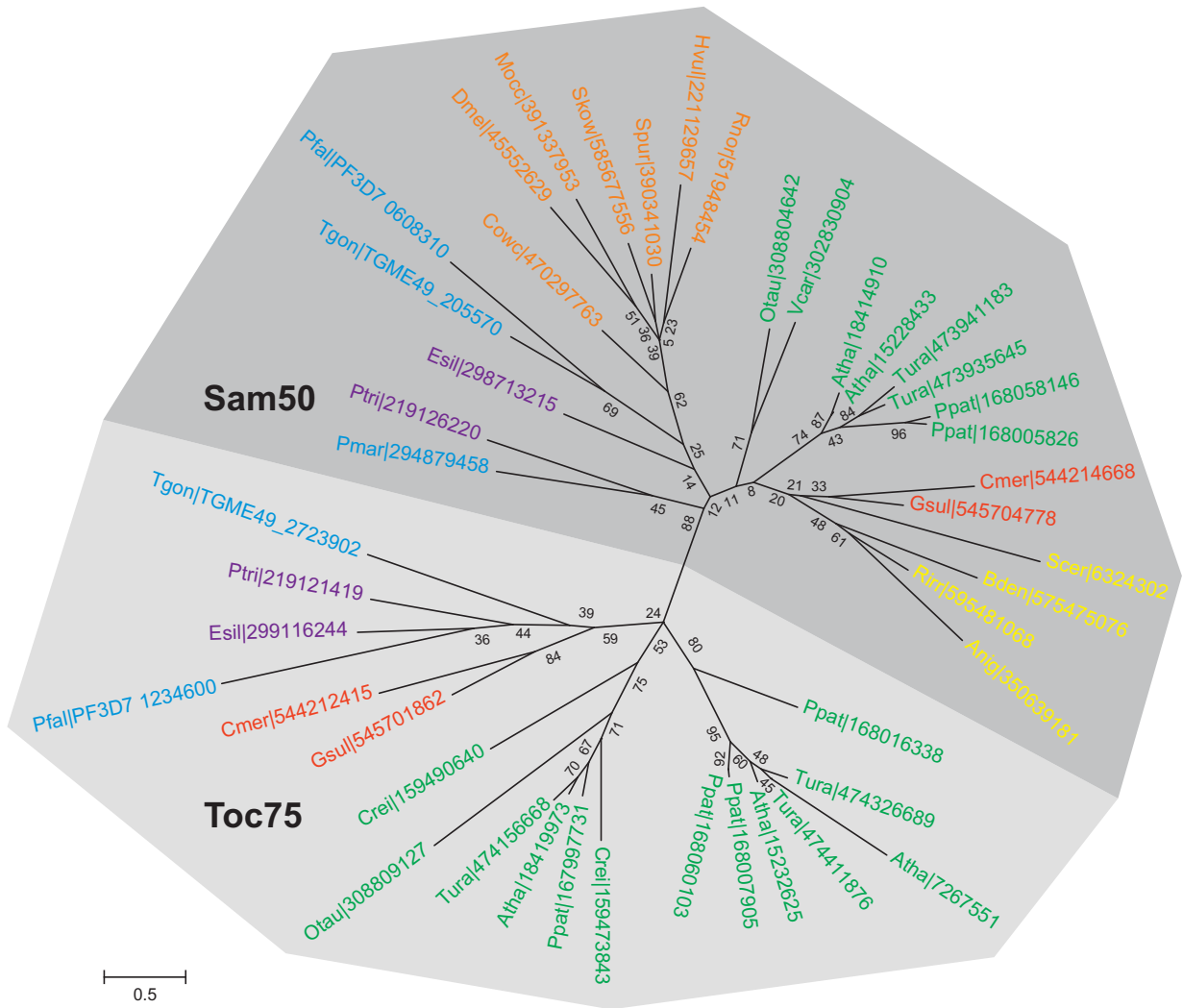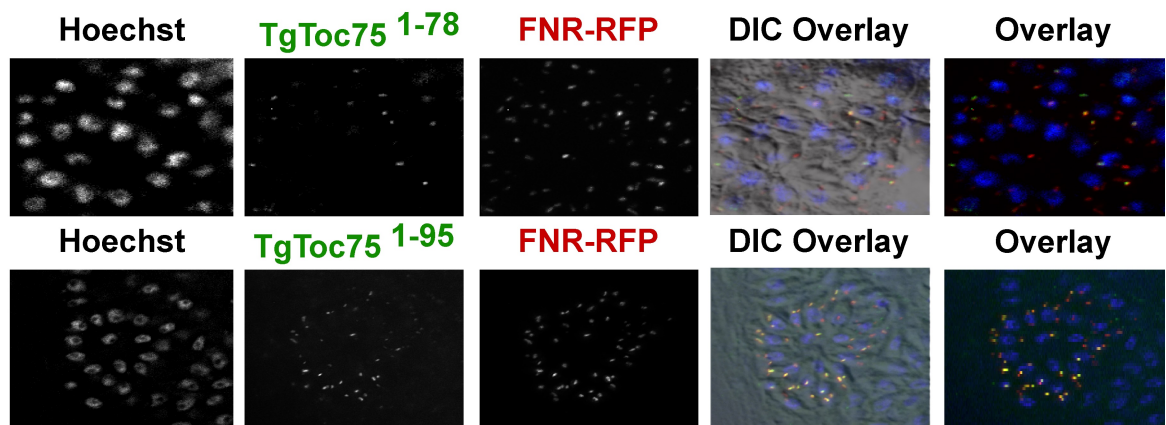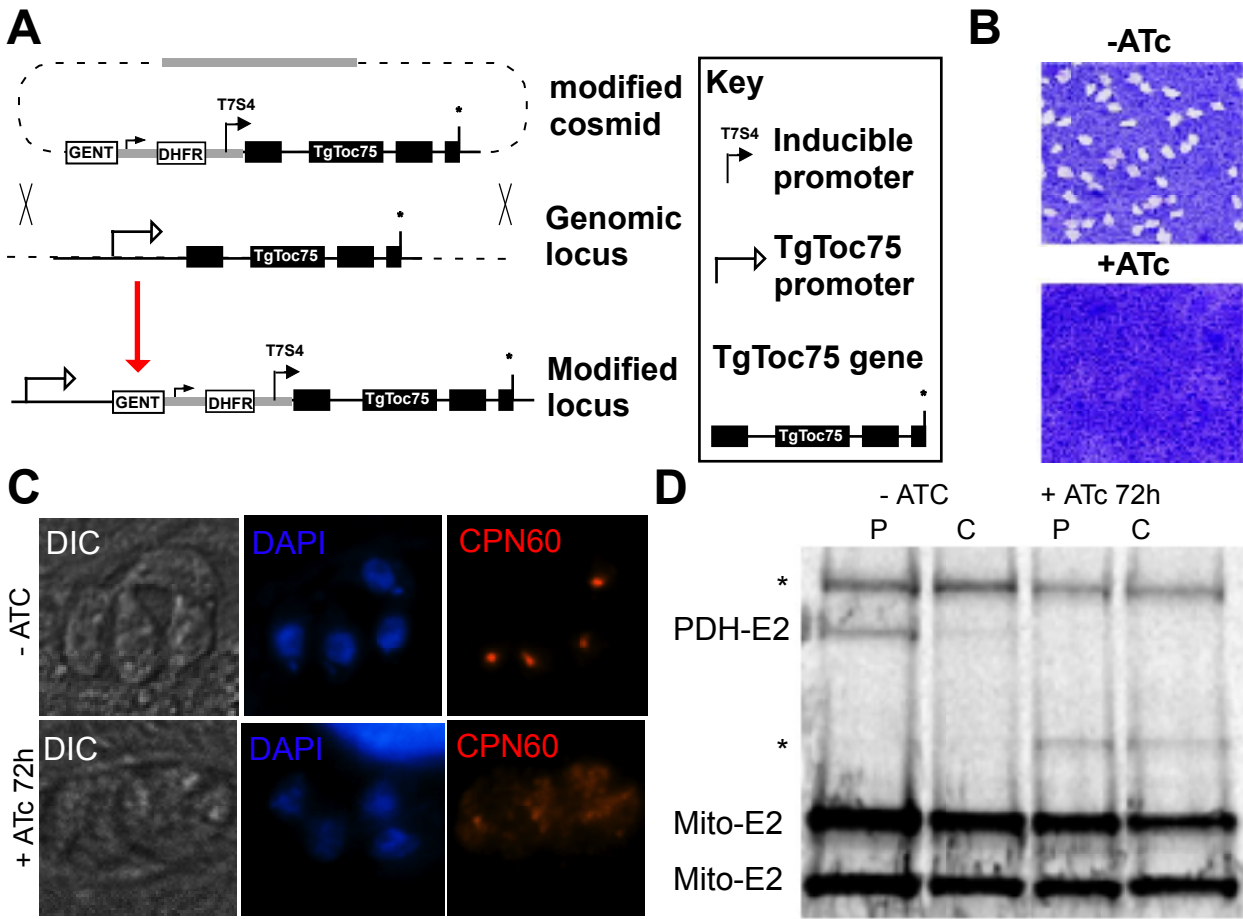
Table 1

	Toc75	SignalP	PlasmoAP	PATS	Sam50	MitoProt	PlasMit
Toxoplasma gondii	*TGME49_072390	N	N	N	TGME49_205570	N	N
Neospora caninum	NCLIV_034910	N	N	N	NCLIV_020120	N	Y
Eimeria falciformis	EfaB_MINUS_25052.g2122	N	N	N	NF	-	-
Eimeria praecox	EPH_0025670	Y	N	N	NF	-	-
Eimeria necatrix	ENH_00027930	N	N	n	ENH_00075630	N	N
Plasmodium falciparum	PF3D7_1234600	Y	Y	Y	*PFF0410w	N	Y
Plasmodium_chabaudi	PCHAS_145150	N	Y	Y	PCHAS_010750	N	Y
Plasmodium_berghei	PBANKA_144920	N	Y	Y	PBANKA_010690	N	Y
Plasmodium_yoelii	PY17X_1451700	N	Y	Y	PY17X_0108400	N	Y
Plasmodium_cynomolgi	PCYB_146020	N	N	N	PCYB_114820	N	Y
Plasmodium_knowlesi	PKH_145170	Y	Y	N	PKH_114100	N	Y
Plasmodium_vivax	PVX_100680	Y	Y	Y	PVX_113574	N	N
Theileria_equi	NF	-	-	-	BEWA_051860	N	N
Babesia_bovis	NF	-	-	-	BBOV_III000300	N	N

* Newer gene model does not agree with our experimental data. See GeneBank accession numbers: TgToc75 KT271755, PfSam50 KT271756

Table S1

Localization	Name	GeneID	SignalP4.1						SignalP3.0					
			C-Score	Y-Score	S-Score	D-Score	SP?	AA at +1	C-Score	Y-Score	S-Score (max)	D-Score	SP?	AA at +1
PPC	PPP1	TGME49_287270	0.11	0.115	0.207	0.123	NO		0.043	0.056	0.269	0.103	NO	
PPC	ATrx2	TGME49_310770	0.198	0.303	0.709	0.369	YES	A	0.299	0.449	0.985	0.665	YES	A
PPC	TgApicE2	TGME49_295990	0.356	0.251	0.405	0.224	NO		0.294	0.405	0.993	0.500	YES	S
PPC	CDC48AP	TGME49_321640	0.356	0.251	0.405	0.224	NO		0.288	0.410	0.944	0.433	YES	V
PPC	TgApicE1	TGME49_314890	0.353	0.188	0.133	0.143	NO		0.200	0.033	0.165	0.089	NO	
PPC	Ubiquitin	TGME49_223125	0.137	0.145	0.215	0.149	NO		0.105	0.085	0.498	0.155	NO	
PPC	UDF1AP	TGME49_285700	0.128	0.136	0.195	0.137	NO		0.067	0.064	0.371	0.100	NO	
outermostmembrane	FtsH	TGME49_259260	0.112	0.12	0.147	0.124	NO		0.087	0.042	0.177	0.069	NO	
outermostmembrane	TGME49_201270	TGME49_201270	0.111	0.106	0.111	0.102	NO		0.034	0.017	0.060	0.017	NO	
outermostmembrane	ATrx1	TGME49_312110	0.265	0.371	0.734	0.418	YES	V	0.370	0.490	0.890	0.468	YES	V
outermostmembrane	APT	TGME49_261070	0.108	0.105	0.157	0.101	NO		0.063	0.044	0.306	0.074	NO	
second inner-most	Toc75/Omp85	TGME49_272390	0.209	0.242	0.685	0.352	YES	V	0.444	0.326	0.993	0.453	YES	V
Innermost	Tic20	TGME49_255370	0.611	0.753	0.99	0.849	YES	Q	0.633	0.701	0.989	0.815	YES	T
Innermost	Tic22	TGME49_286050	0.694	0.58	0.691	0.519	YES	L	0.740	0.755	0.919	0.735	YES	L
luminal	PDH E1b	TGME49_272290	0.415	0.608	0.963	0.77	YES	S	0.539	0.442	0.967	0.653	YES	L
luminal	FabH	TGME49_231890	0.118	0.121	0.157	0.119	NO		0.072	0.073	0.409	0.124	NO	
luminal	PDH E2	TGME49_206610	0.524	0.662	0.961	0.754	YES	Q	0.599	0.625	0.961	0.740	YES	L
luminal	PGKII	TGME49_225990	0.129	0.15	0.279	0.149	NO		0.124	0.080	0.846	0.127	NO	
luminal	FabD	TGME49_225990	0.379	0.37	0.53	0.391	YES	P	0.532	0.530	0.885	0.586	YES	V
luminal	PDH E3	TGME49_305980	0.125	0.23	0.56	0.312	NO		0.127	0.192	0.924	0.495	YES	R
luminal	GyraseB	TGME49_297780	0.109	0.117	0.137	0.117	NO		0.030	0.039	0.310	0.130	NO	
luminal	GyraseA	TGME49_221330	0.609	0.693	0.956	0.763	YES	R	0.679	0.614	0.927	0.069	YES	R
luminal	FNR	TGME49_298990	0.198	0.262	0.641	0.321	NO		0.204	0.358	0.950	0.600	YES	V
luminal	ACP	TGME49_264080	0.47	0.629	0.929	0.744	YES	Y	0.628	0.678	0.970	0.715	YES	F
luminal	TGME49_239680	TGME49_239680	0.275	0.286	0.382	0.291	NO		0.647	0.379	0.985	0.525	YES	S
luminal	TPI-II	TGME49_233500	0.437	0.46	0.765	0.489	YES	F	0.236	0.320	0.913	0.401	YES	F
luminal	FabI	TGME49_251930	0.271	0.399	0.914	0.53	YES	F	0.470	0.402	0.974	0.522	YES	F
luminal	PYKII	TGME49_299070	0.12	0.112	0.132	0.114	NO		0.438	0.079	0.110	0.058	NO	
luminal	RP128	TGME49_209710	0.558	0.477	0.7	0.474	YES	F	0.937	0.707	0.958	0.753	YES	F
luminal	CPN60	TGME49_240600	0.139	0.192	0.349	0.224	NO		0.259	0.359	0.861	0.465	YES	F
luminal	ACC1	TGME49_221320	0.221	0.276	0.52	0.313	NO		0.533	0.446	0.748	0.493	YES	A
luminal	PDH Ea	TGME49_245670	0.128	0.151	0.254	0.148	NO		0.274	0.137	0.876	0.213	YES	P
luminal	ICDH2	TGME49_266760	0.822	0.736	0.892	0.699	YES	S	0.928	0.711	0.901	0.709	YES	R
luminal	YbaK	TGME49_255680	0.107	0.102	0.111	0.099	NO		0.027	0.022	0.072	0.039	NO	
luminal	UROD	TGME49_289940	0.116	0.114	0.163	0.111	NO		0.102	0.139	0.346	0.138	NO	
luminal	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase domain-containing protein	TGME49_055690	0.503	0.653	0.944	0.654	YES	P	0.352	0.501	0.943	0.576	YES	S
luminal	Product: 1-deoxy-D-xylulose-5-phosphate synthase	TGME49_008820	0.11	0.107	0.13	0.106	NO		0.062	0.033	0.119	0.035	NO	
luminal	HU	TGME49_027970	0.235	0.468	0.968	0.72	YES	E	0.364	0.529	0.995	0.700	YES	E
luminal	LipB	TGME49_115640	0.14	0.145	0.34	0.192	NO		0.175	0.116	0.412	0.128	NO	
luminal	RNA helicase	TGME49_291670	0.387	0.497	0.95	0.678	YES	L	0.615	0.572	0.979	0.714	YES	V
luminal	hypothetical	TGME49_059230	0.708	0.556	0.712	0.467	YES	R	0.458	0.626	0.986	0.728	YES	H
luminal	hypothetical	TGME49_039320	0.112	0.132	0.207	0.143	NO		0.052	0.060	0.617	0.172	NO	
luminal	hypothetical	TGME49_039680	0.275	0.286	0.382	0.29	NO		0.647	0.379	0.985	0.525	YES	S
luminal	hypothetical	TGME49_002440	0.178	0.239	0.412	0.24	NO		0.126	0.296	0.947	0.391	YES	W
luminal	hypothetical	TGME49_001270	0.111	0.106	0.111	0.102	NO		0.034	0.017	0.060	0.017	NO	
luminal	NFU	TGME49_021920	0.148	0.222	0.482	0.218	NO		0.415	0.282	0.902	0.338	YES	H
luminal	DNAdDNAh	TGME49_008840	0.111	0.11	0.134	0.104	NO		0.055	0.054	0.527	0.291	NO	

Table S2 - primers used in this study

Primer name	Primer sequence	Purpose
TgToc75_EcoRI_F	CCGAATTCATGGCGGAGGAAGAAAGAC	Forward to amplify TgToc75[78] for ectopic expression in *Toxoplasma*
TgToc75_78_NsiI_R	CCATGCATAGAAACTGGAGAAGACCC	Reverse to amplify TgToc75[78] for ectopic expression in *Toxoplasma*
TgToc75_95_NsiI_R	CCATGCATAAGAGGGGCGGGGGTGC	Reverse to amplify TgToc75[95] for ectopic expression in *Toxoplasma*
TgToc75_277_NsiI_R	CCATGCATTCACGATATCCACGAAGGTACG	Reverse to amplify TgToc75[277] for ectopic expression in *Toxoplasma*
TgToc75_512_NsiI_R	CCATGCATAAACTGCGTCGTCTGTCGTCTG	Reverse to amplify TgToc75[512] for ectopic expression in *Toxoplasma*
TgToc75_790_NsiI_R	CCATGCATAGCCTGCGAACGACGACGCCTC	Reverse to amplify TgToc75[790] for ectopic expression in *Toxoplasma*
TgToc75_FL_NsiI_R	CCATGCATTGAAGCTGTTGTCGGCCACG	Reverse to amplify TgToc75[full-HA]/[Ty] for ectopic expression in *Toxoplasma*
TgSam50_EcoRI_F	CCGAATTCATGGCGGGGTCAGCTCC	Forward to amplify TgSam50[full-HA] for ectopic expression in *Toxoplasma*
TgSam50_NsiI_R	GGATGCATACTACTCGGGGAGTCTTCC	Forward to amplify TgSam50[full-HA] for ectopic expression in *Toxoplasma*
PfOToc75_X_F	AACTCGAGATGAAAAATGTTTTAAGAAAATATAC	Forward to amplify PfToc75[78] for ectopic expression in *Plasmodium*
PfToc75_78_A_R	GGCCTAGGTCTTGTTGTTAGCTTATTCCATAATTC	Reverse to amplify PfToc75[78] for ectopic expression in *Plasmodium*
PfSam50_X_F	CTCGAGATGTTTAATTATTTTTTAAGAAGC	Forward to amplify PfSam50[N] for ectopic expression in *Plasmodium*
PfSam50_60_A_R	AACCTAGGTAAACAAAAATGCTTCCAAAATAATGG	Reverse to amplify PfSam50[N] for ectopic expression in *Plasmodium*
PfOmp85_95_A_R	GGCCTAGGTCCTGTTTCTTCATTTTCTGTTTCC	Reverse to amplify PfToc[95] for ectopic expression in *Plasmodium*
Toc75prorepcosf	GTATGCACATGTCTCTTTTCTGAATCTTTCGCATGAGAAG CAATGCTCCATCGAATGGTAACCGACAAACGCGTTC	Cosmid recombineering to create promoter replacement vector.
Toc75prorepcosr	AGTCCACGACTCAAAAGAGCGAAACGTGTGTTTCTACGGT CGCTCAACGTAGATCTGGTTGAAGACAGACGAAAGC	Cosmid recombineering to create promoter replacement vector.
toc75cosproinserf	ACGTTGAGCGACCGTAGAAACACACGTTTCGCTCTTTTGA GTCGTGGACTGAATGGTAACCGACAAACGCGTTC	Cosmid recombineering to create promoter insertion vector.
toc75cosprionsrev	ATTGAACACCGCCGCCGTGGCGACGATGCCTGTCTTTCTT CCTCCGCCATTTTAGATCTGGTTGAAGACAGACGAAAGC	Cosmid recombineering to create promoter insertion vector.
HA_NsiI_F	CCATGCATTACCCGTACGAC	Primer to amplify 3xHA tag
HA_PacI_R	GGTTAATTAATTAGAGCTCGGC	Primer to amplify 3xHA tag
Apg-qPCR-F	TCTATTGCAATGGAAAAAGGTATG	qPCR to score apicoplast genome
Apg-qPCR-R	TCAATGGTAGAGCAAAGGACTG	qPCR to score apicoplast genome
UPRT-qPCR-F	ACTGCGACGACATACTGGAGAAC	qPCR to score nuclear genome
UPRT-qPCR-R	AAGAAAACAAAGCGGAACAACAA	qPCR to score nuclear genome
ACP$_{F/A}$mutF	*CTGATCAGGCCTGGTGACACA*AGC*ACCGTAGGAAGAAGCAA TGG*	Mutagenesis of F at position +1 of ACP to A
ACP$_{F/A}$mutR	*CCATTGCTTCTTCCTACGGT*GCT*GTGTCACCAGGCCTGAT CAG*	Mutagenesis of F at position +1 of ACP to A
ACP$_{Y/A}$mutF	*CCTGGTGACACAAAAC*CGG*CGGAAGAAGCAATGGATG*	Mutagenesis of Y at alternative position +1 of ACP to A
ACP$_{Y/A}$mutR	*CAT*CCATTGCTTCTTCC*GCC*GGTTTTGTGTCACCAGG*	Mutagenesis of Y at alternative position +1 of ACP to A